

Essays on nonstationary nonlinear time series models

Citation for published version (APA):

Lin, Y. (2021). *Essays on nonstationary nonlinear time series models*. [Doctoral Thesis, Maastricht University]. Global Academic Press. <https://doi.org/10.26481/dis.20211006yl>

Document status and date:

Published: 01/01/2021

DOI:

[10.26481/dis.20211006yl](https://doi.org/10.26481/dis.20211006yl)

Document Version:

Publisher's PDF, also known as Version of record

Please check the document version of this publication:

- A submitted manuscript is the version of the article upon submission and before peer-review. There can be important differences between the submitted version and the official published version of record. People interested in the research are advised to contact the author for the final version of the publication, or visit the DOI to the publisher's website.
- The final author version and the galley proof are versions of the publication after peer review.
- The final published version features the final layout of the paper including the volume, issue and page numbers.

[Link to publication](#)

General rights

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal.

If the publication is distributed under the terms of Article 25fa of the Dutch Copyright Act, indicated by the "Taverne" license above, please follow below link for the End User Agreement:

www.umlib.nl/taverne-license

Take down policy

If you believe that this document breaches copyright please contact us at:

repository@maastrichtuniversity.nl

providing details and we will investigate your claim.

Doctoral thesis

**ESSAYS ON NONSTATIONARY
NONLINEAR TIME SERIES MODELS**

Yicong Lin

2021

ESSAYS ON NONSTATIONARY NONLINEAR TIME SERIES MODELS

Dissertation

To obtain the degree of Doctor at Maastricht University,
on the authority of the Rector Magnificus, Prof. Dr. R.M. Letschert,
in accordance with the decision of the Board of Deans,
to be defended in public
on Wednesday, 6 October 2021, at 10.00 hours

by

Yicong Lin

Promotors

Dr. Stephan Smeekes

Prof. Dr. Eric Beutner, Vrije Universiteit Amsterdam, the Netherlands

Assessment Committee

Prof. Dr. Alain Hecq (Chair)

Prof. Dr. Christoph Hanck, Universität Duisburg-Essen, Germany

Prof. Dr. Robert Taylor, University of Essex, the United Kingdom

Dr. Ines Wilms

© Yicong Lin, Maastricht 2021.

All rights reserved. No part of this publication may be reproduced, stored in a retrieval system or transmitted in any form or by any means, electronic, mechanical, photocopying, recording or otherwise, without prior written permission of the author.

Cover Woodcut by Yik Wong, 2021

Publisher Global Academic Press

ISBN 978-94-6423-445-9

Printing ProefschriftMaken, the Netherlands

Dedicated to the memory of Jean-Pierre Urbain

Acknowledgments

My parents still do not understand what I am working on, but they approximated it quite well. The Chinese character for “Yi” in my name means “more than”, and the one for “Cong” means “being smart”. Putting together, “Yi-Cong” gives an approximation to “a wise person” by setting the lower bound “more than just being smart”. You see, these two characters already reflect much of my work; it is all about bounds and approximations. But if I tell you this is why I do econometrics, I would be just joking. Working on econometrics was not part of my original plan. It was a decision made after many talks with my dearest professor Jean-Pierre Urbain under the trees in the schoolyard, and possibly the mental stimulation of his second-hand smoke. I never regret the decision. Instead, I am grateful to Jean-Pierre for bringing me to this field and supporting me in various aspects in my first two years in Maastricht. After his sudden death, day by day, year by year, I still often could not help looking at the place where he stood when I passed by – those trees, the man with a cigarette, getting closer, then those smiles.

My Ph.D. career would not go well without the excellent guidance of my daily supervisors Eric Beutner and Stephan Smeekes. I am deeply indebted to both of you for your kindness and patience, and for giving me much freedom to do my research and come up with my ideas. Many thanks for always being approachable and never give equivocal answers whenever I have questions. The ways you teach and do research have set me the best examples as an educator and researcher in my career. Further, Eric, I admire your rigorousness. Although there is “quasi-Hadamard (differentiability)” to you, there is never “quasi-Beutner (rigorousness)” to me. Thank you also for the short, snappy counter-examples. They are more intense than double espresso shots before sleep, giving me the most sleepless nights if you are not the “rock star” living upstairs. Stephan, I appreciate your many insightful remarks made in different

occasions. They are often like spring's honied cud that I love to ruminate, especially when Eric serves too much espresso.

There would be no this section without the positive assessment of the reading committee. I therefore would like to thank Christoph Hanck, Alain Hecq, Robert Taylor, and Ines Wilms, for their time and effort to evaluate and the approval of my thesis manuscript. I am also grateful to Xiaohu Wang for hosting me for three months at The Chinese University of Hong Kong. Thank you for frequent personal calls and many valuable discussions on new ideas up until now. Finally, I express my gratitude to all of my colleagues from the Department of Quantitative Economics. I want to thank two people in particular: Dries Vermeulen for his enjoyable course in measure theory and topology; Michael Eichler for whom I was tutoring the Probability Theory course for three years. I benefited greatly from both of you.

On a more personal level, let me first send my appreciation to Caterina and Etiënne, not only for their cuisines but also for their support as a friend. I would especially like to thank Caterina for telling me what is good but not right – it is good to have a pizza with pineapple and chicken on the top but it is not right. It saves my life when I know more Italians later on. I also would like to thank my office mates Francesco and Roland for co-creating the sleepy afternoons – after coffee breaks. We will soonish be Oscar-nominated for the best acting in “Work Hard”. All jokes aside, sharing an office and further becoming good friends with you were greatly joyful. Let me also send my appreciation to some old Maastricht people in the current VU group after my Ph.D. contract: Hande, Marina, and Sean. With you, the starting (completely virtual) year becomes much easier to pull through.

Four seasons fill the measure of the year; there is Ginger – a Chinese restaurant near the school – filling my four seasons. Thank you, Faye and Jian (from Ginger), my thesis was written in the ink made of your soy sauce and yellow curry sauce. I also feel blessed to be close to a small part of the Chinese community in Maastricht: Honglin Chen, Bin Dong, Jianan Hou, Ming Li, Runliang Li, Shaokang Li, Fan Rao, Li Sun. In particular, Li, your happiness

and positive life attitude are contagious to me when I am down. Jianan and Ming, juicy news never travel faster than your messages. My Ph.D. life would taste duller without your spice. In addition, I want to thank my high school bff (best friend forever) Sophia Mok and her siblings, living in Belgium currently. Thank you for visiting me and bringing me delicious handmade Tiramisu – the body weight – from time to time.

Now I will turn to the multifunctional (book/coffee/food/music) group Hanno and Yuliya, without whom my Ph.D. could have been finished much earlier if I have not spent so much time on enjoying a cup of coffee, taking walks, and making bad jokes. As for Hanno, I especially thank the proper distance between Maastricht and Rotterdam, preventing us from killing each other over minor things while writing our first coauthored paper (Chapter 3). On the other hand, I cannot miss you more when the physical distance separating us for too long and cannot laugh harder whenever you are around. Also thank you for introducing me to your family, and always being like an older, different-looking (thank God) brother. Yuliya, thank you for inviting me mountain-biking in Buxton, without providing enough food for energy, from which I understand hunger and learn the importance of life insurance. I also thank you for not yet inviting me to walk on a tightrope over a cliff, otherwise my thesis would certainly cease to exist. Nonetheless, those exciting and heart-warming moments with you and your smiles make up a sunny spring in the tranquil cove of my heart.

There is also a group of friends backing me up in different ways, thousands of miles away. I am grateful to Huifang Chen, Chujun Hu, Shuxian Huang, Tingting Huang, Guangzhao Liang, Weibin Wang, Yik Wong, Jiangying Yu, Bowen Zhang, and Guicong Zhang, for their deep friendship since even childhood. I thank Yik chai (“buddy” in Cantonese) for making the woodcut for my thesis cover. Moreover, I would like to address a special word of thanks to Bin Lin, Mimi Qimei Huang, and Dixiu Yang. You are my chosen family. I am also grateful to have two lovable godsons An’an Anrui and AJ Zhengxi in the family!

Like wild horses, I enjoy the freedom and warmth of the sun. But no matter

where I gallop to, I know I will be homeward bound. I consider myself extremely lucky that I was born in a large but close-knit family. Many thanks to my cousin Danxia for constantly mailing me books, hometown specialties and our tea in the past years. I am deeply grateful to my cousins Lili, Lin (Lin), Liying, Weitao, and Jimmy Yatchung, for partly playing my role as the child to my parents to fill my absence. I dedicate the very last sentences to my parents for their unconditional love, for those who never give up on me. You brought me home when I ran away from it at the age of thirteen, holding me back when I was bordering on break laws. Today every honor of mine also belongs to you. Persimmons have ripen, let's pick them home.

Yicong Lin
Bergen, Norway
August 2021

Contents

Acknowledgments	vii
1 Introduction	1
1.1 Nonlinearity	2
1.2 Trends and the Models	3
1.3 The Modified Cholesky Decomposition (MCD)	6
1.4 Outline	9
2 GLS Estimation and Confidence Sets for the Date of a Single Break in Models with Trends	11
2.1 Introduction	13
2.2 The Model and Infeasible GLS	15
2.3 Asymptotic Theory	19
2.4 Confidence Sets for Break Dates	30
2.5 Prewhitening to Improve Finite-Sample Properties	33
2.6 Simulations	36
2.7 Conclusion	41
Appendix A Preliminary Results for Infeasible GLS	42
Appendix B Infeasible Generalized Least Squares	64
Appendix C Consistency of Estimated BIAM	71
Appendix D Feasible Generalized Least Squares	78
Appendix E A Hájek and Rényi Inequality	83
Appendix F Additional Simulation Outputs	84
3 Fully Modified GLS with an Application to the Environmental Kuznets Curve	89
3.1 Introduction	91
3.2 The Model	93
3.3 Asymptotic Theory	97

3.4	Simulations	109
3.5	Empirical Application	120
3.6	Conclusion	125
Appendix A	Auxiliary Results	126
Appendix B	Proofs of Main Theorems	135
Appendix C	Multivariate KPSS-type Tests	144
Appendix D	Estimating Quantities for FM Inference	145
Appendix E	Details on Implementation	147
Appendix F	Additional Information for EKC	148
4	Cointegrating Polynomial Regressions with Power Law Trends: Environmental Kuznets Curve or Omitted Time Effects?	151
4.1	Introduction	153
4.2	The Model and NLS Estimation	158
4.3	Asymptotic Theory	160
4.4	Simulations	170
4.5	Empirical Application	175
4.6	Summary and conclusion	181
Appendix A	Useful Lemmas	182
Appendix B	Proof of Main Theorems	197
Appendix C	Additional Theoretical Results	210
Appendix D	Simulation Results on Confidence Intervals for θ	216
Appendix E	Further Empirical Results	217
5	Conclusion	227
	Bibliography	231
	Impact Paragraph	247
	About the author	251

1

Introduction

This thesis focuses on modeling economic relationships in the framework of time series analysis. We will discuss three different topics in the core chapters. The reader may read these chapters in an arbitrary order because there is no linear storyline. In fact, ‘nonlinearity’ itself is the main theme. Namely, we will consider some specific models of nonlinear functional forms throughout the thesis as the first step of a long journey; a journey to comprehend complex, intrinsically nonlinear economic relationships.

Up to this point, you may have a number of Socratic questions: What is a nonlinear model, and why is it important? A clear definition of nonlinearity is needed before starting our journey.

1.1 Nonlinearity

We understand the (non)linearity in the sense of Lee et al. (1993). More specifically, let y_t be an observation on the dependent variable at time t and \mathbf{x}_t be a $d \times 1$ vector containing our (economic) variables of interest. In general, \mathbf{x}_t may but need not necessarily contain a constant and lags of y_t . We say the process $\{y_t\}$ is not linear in mean conditional on \mathbf{x}_t if

$$\mathbb{P}[\mathbb{E}(y_t | \mathbf{x}_t) = \mathbf{x}_t' \boldsymbol{\gamma}] < 1, \quad \text{for all } \boldsymbol{\gamma} \in \mathbb{R}^d. \quad (1.1)$$

If for some $\boldsymbol{\gamma} \in \mathbb{R}^d$, the probability above is 1, then it is said to be linear.

Why nonlinearity? One may largely agree that linear models are often useful approximations to complicated economic relationships. In some cases, nonlinear relations refine existing economic/econometric theory, originating from empirical evidence. For instance, presuming that economic variables are jointly Gaussian justifies the use of linear models. However, considerable evidence shows its inconsistency with economic data (see e.g. Dorfman (1993), Richardson and Smith (1993), Dufour et al. (1998) and references therein). There are many causes. As has been pointed out by Teräsvirta et al. (2010),

“Any nonlinearity will arise from the structure and institutions of the economy: its rules and laws, the behaviour of its agents and corporations, and, possibly, of its policy makers.”

In other cases, the economic theory postulates a nonlinear relationship from the outset. Simple examples include Phillips curve, production and supply functions, and saving functions. Other examples such as disequilibrium models and labor market models can be found in Chapter 2 of Teräsvirta et al. (2010).

1.2 Trends and the Models

When nonlinearity is allowed, a natural question arises: among (infinitely) many choices, which model/approach to use? This thesis will focus on *discrete* time series models

$$y_t = f(x_t, \gamma) + u_t, \quad t = 1, 2, \dots, T, \quad (1.2)$$

where f is some *known* function, u_t is an unobservable disturbance, and γ is an unknown parameter vector. As will be discussed later, we are interested in economic variables that contain a trend. In time series econometrics, trending variables are not stationary – the property saying certain statistical properties of the process do not change over time. We will focus on two commonly adopted trends: deterministic ($x_t = t$) and stochastic ($x_t - x_{t-1} = v_t$, where v_t is stationary). For a much wider class of nonlinear stochastic trends, please refer to Granger et al. (1997). Furthermore, we will allow for the error-regressor correlations (endogeneity) whenever the regressors are stochastic. The main goal of the thesis is to develop estimation and inference methods for the unknown parameter vector γ . Now, we briefly motivate and introduce the models studied in the main chapters.

There are mainly two types of models considered. The first type, which is referred to as *trend break model*, stems from some empirical evidence in different disciplines. Researchers often employ a linear time trend to describe

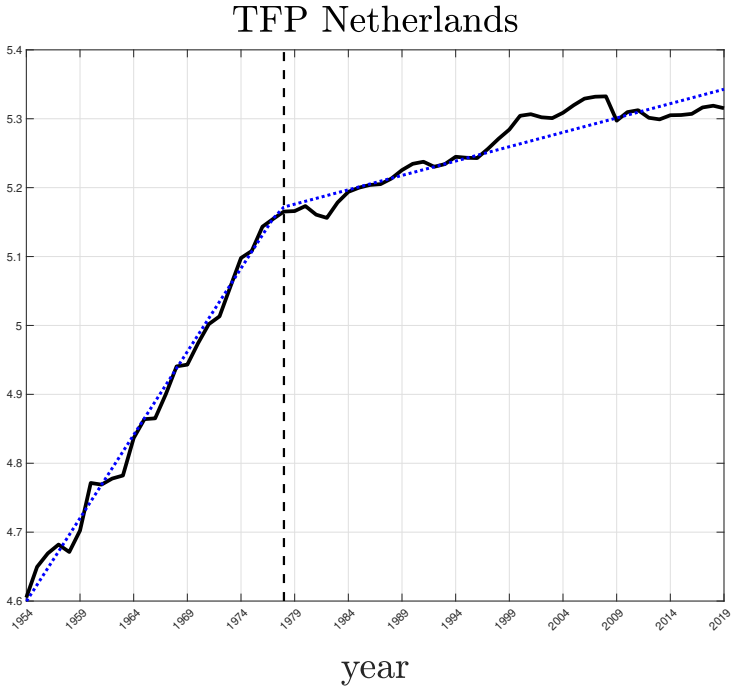


Figure 1.1: Total factor productivity at constant national prices for the Netherlands (1954 = 100, in log, solid line). Data are available on FRED, prepared by Feenstra et al. (2015). Using the method proposed in Chapter 2, the estimated location of change is in year 1978.

the long-term, smooth tendencies that appear in the levels of data. In many situations, the estimation of parameters in time series models often presumes that the parameters do not change over time. Due to various factors such as policy interventions, parameter stability is often rejected in empirical studies. The inference for parameters in the trend break model is thus of importance (Harvey et al. (2009)). Moreover, the exact locations of changes are usually unknown. The identification of them is not only crucial, for instance, to analyze the effects of a policy change, but also for forecasting performance (Pesaran and Timmermann (2005, 2007), Boot and Pick (2020), Rossi (2021)). As an illustration, we display the total factor productivity (TFP) for the Netherlands (1954-2019) in Figure 1.1. There are essentially two phases of the TFP growth in the Netherlands, namely 1954-1978 and 1978-2019. The growth patterns

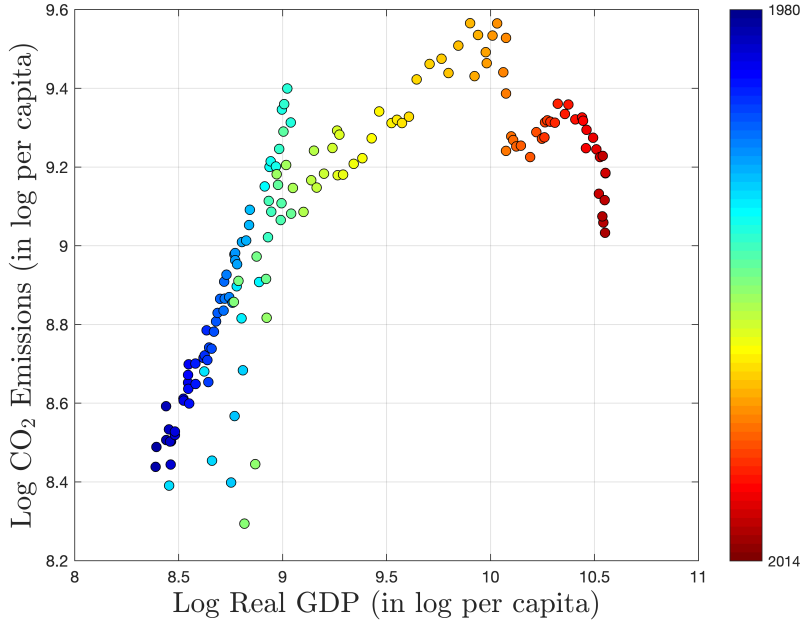


Figure 1.2: The log carbon dioxide emissions per capita versus the log GDP per capita for Belgium, adopted from Chapter 4

in both phases are smooth, but there is a slowdown in the second period. Such nonlinear behavior reminds us of the long-standing discussion among economists about economic slowdown, see e.g., Krugman (1995), Gordon (2016), Banerjee and Duflo (2019). To model it, we will consider trend break models as mentioned, see Chapter 2.

If a nonlinear relationship is suggested by the economic theory, one may have some guidance as to the functional form of f . Simon Kuznets (Nobel Prize winner 1971) has described an inverse U-shaped relationship between per capita income and income inequality, nowadays known as the Kuznets curve. This concept is later introduced in environmental economics by Grossman and Krueger (1995) as the environmental Kuznets curve (EKC), becoming a popular area of research with increasing attention to the ecological environment in recent years. The EKC postulates a similar inverse U-shaped relationship

between per capita income and environmental degradation (Stern (2004)). An example of Belgium (1870-2014) is shown in Figure 1.2.

This conjecture motivates a polynomial representation of f which will be considered as the second type of models in the thesis. If a regressor x_t is stationary, its integer powers are also stationary. In this case, the inference on unknown parameters is not much different from classical cases. However, a different framework is needed because the regressor, which is the log GDP per capita the EKC, is widely believed to be nonstationary, see Chapter 3 and Chapter 4.

Of course, many other models are useful in practical studies. However, they fall outside the scope of the thesis. These include large classes of, for instance, smooth transition models (Teräsvirta et al. (2010), Chapter 3), time-varying models (Zhang and Wu (2015), Li et al. (2020)) and transformation models (Lin et al. (2020) and references therein). Recent machine learning techniques for nonlinear models will also not be covered such as neural network models and boosting methods (Phillips and Shi (2021)). Finally, we will not consider the chaos theory. Although deterministic, chaotic dynamics are often used to model nonlinearity in many physical sciences, economic variables are generally not chaotic (Liu et al. (1992)). We refer interested readers to the monograph by Kantz and Schreiber (2003).

1.3 The Modified Cholesky Decomposition (MCD)

Although our models are nonlinear in regressors, as will be seen later, they are linear in parameters sometimes. In a similar spirit of the Gauss-Markov Theorem (Kariya and Kurata (2004), Theorem 2.1), we will consider the *generalized least squares* (GLS) estimation to improve the accuracy of estimators in small samples. In time series, GLS estimators are typically constructed using the $T \times T$ inverse autocovariance matrix Σ^{-1} , where

$$\Sigma = \mathbb{E}(\mathbf{u}\mathbf{u}'), \quad \mathbf{u} = (u_1, u_2, \dots, u_T)'. \quad (1.3)$$

Since the error process $\{u_t\}$ is unobservable, Σ and, more importantly, Σ^{-1} are unknown and have to be estimated in practice.

In general, Σ does not possess a simple structure as is usually assumed in textbooks, e.g. Greene (2012). Traditional methods often fail due to the large dimensionality as T increases. Therefore, many approaches such as banding and tapering sample autocovariance matrices have been proposed to deal with the estimation of Σ , see e.g. Bickel and Gel (2011), Bickel and Levina (2008), McMurtry and Politis (2010), Wu and Pourahmadi (2009), Xiao and Wu (2012) and the book by Pourahmadi (2013). For GLS, it is more important to estimate the inverse matrix Σ^{-1} . To this end, we introduce a matrix decomposition that serves as the basis of our approach (Chapter 2 and Chapter 3): the Modified Cholesky Decomposition (MCD) by Pourahmadi (1999).

Suppose A is a T -dimensional positive definite matrix. Recall the Cholesky decomposition, one can write $A = LDL'$, where L is a lower unit triangular matrix and D is a diagonal matrix. The MCD decomposes A into the same structure. For illustrations, consider a stationary autoregressive process of order 1, namely

$$u_t = \rho u_{t-1} + \epsilon_t,$$

where $|\rho| < 1$ and $\{\epsilon_t\}$ is an independent, zero mean process with variance σ_ϵ^2 . Now we define a lower unit triangular matrix as

$$F = \begin{pmatrix} 1 & & & \\ -\rho & 1 & & \\ & \ddots & \ddots & \\ & & -\rho & 1 \end{pmatrix}, \quad \text{and then } Fu = \begin{pmatrix} u_1 \\ u_2 - \rho u_1 \\ \vdots \\ u_T - \rho u_{T-1} \end{pmatrix} = \begin{pmatrix} u_1 \\ \epsilon_2 \\ \vdots \\ \epsilon_T \end{pmatrix}.$$

By the independence,

$$F \Sigma F' = \mathbb{E}(F u u' F') = \begin{pmatrix} \mathbb{E} u_1^2 & & & \\ & \mathbb{E} \epsilon_2^2 & & \\ & & \ddots & \\ & & & \mathbb{E} \epsilon_T^2 \end{pmatrix} = \sigma_\epsilon^2 \begin{pmatrix} \frac{1}{1-\rho^2} & & & \\ & 1 & & \\ & & \ddots & \\ & & & 1 \end{pmatrix} =: S.$$

We obtain

$$\mathbf{F} \boldsymbol{\Sigma} \mathbf{F}' = \mathbf{S} \quad \text{and} \quad \boldsymbol{\Sigma}^{-1} = \mathbf{F}' \mathbf{S}^{-1} \mathbf{F}. \quad (1.4)$$

The key ingredient of the decomposition above is the orthogonality among the elements of $\{u_1, \epsilon_2, \dots, \epsilon_T\}$ after a linear transformation of $\{u_1, u_2, \dots, u_T\}$. For a general stationary process $\{u_t\}$, the orthogonality can be obtained by linear projections, i.e. the Gram-Schmidt process. The process can be described as follows.

- (i) Project u_2 on the space spanned by u_1 . It can be written as $a_1(1) = \arg \min_{c_1 \in \mathbb{R}} \mathbb{E} (u_2 - c_1 u_1)^2$. Then $u_2 - a_1(1)u_1$ is orthogonal to u_1 .
- (ii) Project u_3 on the space spanned by u_2 and u_1 . Namely, $[a_1(2), a_2(2)]' = \arg \min_{(c_1, c_2)' \in \mathbb{R}^2} \mathbb{E} (u_3 - c_1 u_2 - c_2 u_1)^2$. Then $u_3 - a_1(2)u_2 - a_2(2)u_1$ is orthogonal to both u_1 and u_2 , and thus $u_2 - a_1(1)u_1$.
- (iii) Continue the procedure, an orthogonal sequence can be obtained.

Now, the MCD can be summarized as multiple minimization problems:

$$[a_1(\ell), a_2(\ell), \dots, a_\ell(\ell)]' = \arg \min_{(c_1, \dots, c_\ell)' \in \mathbb{R}^\ell} \mathbb{E} (u_t - c_1 u_{t-1} - \dots - c_\ell u_{t-\ell})^2,$$

where $1 \leq \ell \leq T - 1$. Let $\mathbf{F} = (F_{ij}, 1 \leq i, j \leq T)$ with

$$F_{ij} = \begin{cases} 0, & i < j, \\ 1, & i = j, \\ -a_{i-j}(i-1), & 2 \leq i \leq T, 1 \leq j \leq i-1. \end{cases}$$

And let $\mathbf{S} = \text{diag}(s_0^2, s_1^2, \dots, s_{T-1}^2)$, where $s_0^2 = \mathbb{E}(u_t^2)$ and

$$s_\ell^2 = \mathbb{E} [u_t - a_1(\ell)u_{t-1} - \dots - a_\ell(\ell)u_{t-\ell}]^2, \quad 1 \leq \ell \leq T - 1.$$

Then we can obtain the MCD decomposition of $\boldsymbol{\Sigma}^{-1} = \mathbf{F}' \mathbf{S}^{-1} \mathbf{F}$. The MCD works not only in time series data that have natural orderings, but also in the

data that can be ordered and predetermined, see Kang and Deng (2020).

1.4 Outline

We have broadly discussed the content of the core chapters in the previous pages. One can additionally find an extended abstract and a detailed introduction at the start of each chapter. Here, we provide a summary of the thesis in a bigger picture. As mentioned, this thesis aims to develop estimation and inference methods for unknown parameters in nonlinear, nonstationary time series models, possibly with endogeneity. A GLS framework for the trend break models is developed in Chapter 2. In Chapter 3, we propose a fully modified GLS framework to study multivariate models that include deterministic trends, stochastic trends and integer powers of stochastic trends, allowing for both cross-sectional and serial dependence. We apply this framework to investigate the EKC hypothesis for multiple countries. Chapter 4 poses a question in the EKC debates for which researchers might need to be aware. The framework of the previous chapter will be partly extended. In Chapter 5, we end this dissertation by reviewing both the introduction and the core chapters. A short outlook on further research questions will be provided. The bibliography, impact paragraph and my curriculum vitae are included as final attachments.

2

GLS Estimation and Confidence Sets for the Date of a Single Break in Models with Trends

This chapter is based on the paper Beutner et al. (2021).

Abstract

We develop a Feasible Generalized Least Squares (FGLS) estimator of the date of structural break in models that allow level and/or trend breaks. The estimator is based on a consistent estimation of the T -dimensional inverse autocovariance matrix of errors that are stacked over time. We find a monotone and continuous cubic polynomial transformation of the break date estimates can be approximated by a nonstandard yet nuisance parameter free distribution asymptotically. The new limiting distribution captures the asymmetry and bimodality observed in finite samples. The polynomial transformation itself contains some unknown parameters which can be estimated consistently, and therefore the asymptotic results can immediately be applied for inference with a single, known, set of critical values, requiring no simulation. We consider the confidence intervals/sets for break dates based on both Wald-type tests and by inverting multiple likelihood ratio (LR) tests. A simulation study shows that the proposed FGLS estimator increases the empirical concentration probability in a small neighborhood of the true break date and potentially reduces the mean squared errors. We also find the LR-based confidence intervals/sets have good coverage while maintaining relatively informative length even with highly persistent error dynamics and small break sizes.

2.1 Introduction

Practitioners often make use of a linear time trend to describe the long-term, smooth tendencies that appear in levels of data. It is not surprising that evidence of structural changes in deterministic trends is widely found in different disciplines such as macroeconomics and climatology, see, e.g., Perron (1989), Raj and Slottje (1994), Estrada and Perron (2017) and Friedrich et al. (2020). The dates of breaks are usually unknown and have to be estimated. To link the estimated break point to external events, the break date has to match the event, and therefore a statistical assessment of the accuracy of the break date estimate is essential. To this end, we develop a generalized least squares (GLS) framework for estimation and inference of the single break date in models that allow breaks in the slope of deterministic linear trend and/or the level.

Two break date estimators are specifically relevant to our analysis. Perron and Zhu (2005) constructed the estimator as the minimizer of the sum of squared residuals from ordinary least squares (OLS). Alternatively, Harvey and Leybourne (2014) employed a quasi-differencing (QD) regression, resulting in a more efficient estimator. Both articles have provided the corresponding limiting distributions that deliver good approximations to finite samples. Unfortunately, their representations depend upon a host of nuisance parameters or the exact distribution of errors, and therefore the inference seems not straightforward. A typical remedy to get rid of these dependencies in the literature is to adopt the asymptotic framework whereby the break sizes are shrinking as the sample size increases (Perron (2006)). Deng and Perron (2006) have found that the limiting distributions developed in Bai (1997) and Bai et al. (1998) cannot capture the signals of trend change, leading to a symmetric and unimodal shape. Using these limiting distributions for inference can be misleading. It is because simulation studies, for instance, in Perron and Zhu (2005), Deng and Perron (2006) and Harvey and Leybourne (2014), have reported that the finite sample distributions of least squares estimators are asymmetric and bimodal.

Our GLS estimator, which differs from the QD approach, further improves the finite sample efficiency by including the full T -dimensional inverse au-

tocovariance matrix of disturbances into construction. Intuitively speaking, our estimator takes the complete dynamics of errors into account, while the QD estimator only captures the strongest serial correlation. We also adopt the shrinking asymptotic framework. Our asymptotic results have two advantages compared to previous studies. First, our asymptotics can be readily applied for inference with a single, known, set of critical values, requiring no simulation. Second, the asymptotic distribution captures the asymmetry and bimodality observed in finite samples. The key ingredient of our asymptotic analysis that allows for these results is a modification of the rates at which the level and slope break sizes shrink. It is found that a monotone and continuous cubic polynomial transformation of the break date estimates can be approximated by a well-known distribution that is nuisance parameter free. The transformation itself contains some unknown parameters which are consistently estimable.

The inverse autocovariance matrix is usually unknown in practice. To obtain a feasible GLS estimator, a consistent estimator of the inverse matrix is thus necessary. Based on the Modified Cholesky Decomposition (MCD) proposed by Pourahmadi (1999), we can easily estimate the inverse matrix element-wise. The elements are obtained by multiple autoregressions using the first-stage regression residuals. Similar constructions based on MCD have been adopted in several papers for different purposes, see Cheng et al. (2015), Ing et al. (2016) and Lin and Reuvers (2020b). Nonetheless, the consistency results in the previous studies can not be applied directly in the current context. The main difference is that our first-stage ‘residuals’ are contaminated by break date estimates which have a slow convergent rate (slower than $\text{root-}T$). As a result, one would need to bound the estimation errors differently. Our result of consistency sheds some light on how the inverse autocovariance matrix estimator is affected by estimates in the first stage.

So far, we have confined our attention to the estimation of break date and its sampling distribution. For constructing confidence sets, one common way is by inverting multiple locally best invariant (LBI) tests (Elliott and Müller (2007), Harvey and Leybourne (2015)). It is well reported that the LBI-based

confidence sets have good empirical size properties irrespective of break magnitudes, see discussions in Chang and Perron (2018) and Yamamoto (2018). Nevertheless, the LBI-based confidence sets are often too wide to be informative, particularly, in the presence of strong serial correlations. We construct confidence intervals/sets differently by looking at the likelihood ratio (LR) tests as previously advocated by Hansen (2000) and Eo and Morley (2015). We find that the test statistic is asymptotically pivotal in the shrinking framework. The confidence sets of break dates then can be constructed by inverting multiple LR tests. The LR-based confidence sets are not necessarily intervals. We then obtain intervals by a simple modification of the LR-based sets. As a benchmark, we also provide construction of intervals using Wald-type tests. Our comprehensive simulation studies show some favorable properties of the LR-based intervals, in terms of both coverage and length.

Throughout this chapter, C denotes a generic independent and positive constant, $\mathbf{1}_{\{\cdot\}}$ is an indicator function and $[a]$ denotes the integer part of a . Let $\|\mathbf{x}\|_p = \left(\sum_{i=1}^n |x_i|^p\right)^{1/p}$ be the p -norm of the vector $\mathbf{x} \in \mathbb{R}^n$, and $\|\mathbf{A}\|_p = \max_{\|\mathbf{x}\|_p=1} \|\mathbf{A}\mathbf{x}\|_p$ be the induced norm of a matrix \mathbf{A} . The subscripts are omitted whenever $p = 2$, i.e. $\|\mathbf{x}\| = \left(\sum_{i=1}^n |x_i|^2\right)^{1/2}$ and $\|\mathbf{A}\| = (\lambda_{\max}(\mathbf{A}'\mathbf{A}))^{1/2}$ where $\lambda_{\max}(\cdot)$ denotes the largest eigenvalue. We use the symbol “ $\xrightarrow{\mathcal{D}}$ ” to signify weak convergence, “ $\xrightarrow{\mathcal{P}}$ ” for convergence in probability, and “ $\stackrel{\mathcal{D}}{=}$ ” for equality in distribution. The stochastic order and strict stochastic order relations are indicated by $O_p(\cdot)$ and $o_p(\cdot)$.

2.2 The Model and Infeasible GLS

Consider the following model that allows breaks in both level and trend:

$$y_t = \mu + \beta t + \mu_b \mathbf{1}_{\{t > k\}} + \beta_b (t - k) \mathbf{1}_{\{t > k\}} + \epsilon_t, \quad t = 1, \dots, T, \quad (2.1)$$

where y_t is a scalar observable variable, k is a potential break point, and ϵ_t is the disturbance. The coefficients μ_b and β_b capture the break sizes in the intercept and trend slope, respectively. The trend function $\beta t + \beta_b (t - k) \mathbf{1}_{\{t > k\}}$ is continuous

at date k . To express the model in matrix form, we define T -dimensional vectors \mathbf{y} and $\boldsymbol{\varepsilon}$ by stacking y_t and ϵ_t over time, respectively. Let $\mathbf{X}_k = (\mathbf{x}_{k,1}, \dots, \mathbf{x}_{k,T})'$ be a $T \times 4$ matrix with $\mathbf{x}_{k,t} = (1, t, \mathbf{1}_{\{t > k\}}, (t - k)\mathbf{1}_{\{t > k\}})'$, and $\boldsymbol{\gamma} = (\mu, \beta, \mu_b, \beta_b)'$. Throughout, the true parameters are denoted with a 0 subscript. In particular, k_0 , the unknown true break date, and $\boldsymbol{\gamma}_0 = (\mu_0, \beta_0, \mu_{b,0}, \beta_{b,0})'$. Hence the data-generating process (DGP) is

$$\mathbf{y} = \mathbf{X}_{k_0} \boldsymbol{\gamma}_0 + \boldsymbol{\varepsilon}. \quad (2.2)$$

The omissions of random components in the model imply that the innovation process $\{\epsilon_t\}$ may exhibit serial dependence. The dependence can also be strong in particular for macroeconomic time series. We assume the dependence can be modeled by a stationary AR(∞) process (also see Assumption 2.3 below), that is,

$$A(L)\epsilon_t = \left(1 - \sum_{j=1}^{\infty} a_j L^j\right) \epsilon_t = \eta_t. \quad (2.3)$$

An efficient estimation of the parameters shall account for the serial correlation. It can be achieved by using GLS which relies on the inverse autocovariance matrix $\boldsymbol{\Sigma}^{-1}$, where

$$\boldsymbol{\Sigma} = (\mathbb{E}(\epsilon_s \epsilon_t), 1 \leq s, t \leq T). \quad (2.4)$$

While $\boldsymbol{\Sigma}$ simply contains the autocovariances of the innovations, the structure of $\boldsymbol{\Sigma}^{-1}$ is seemingly complicated. An application of the modified Cholesky decomposition (MCD) developed by Pourahmadi (1999) provides us some insights into GLS. This method relates matrix decompositions with linear minimum mean squared error (MSE) predictors. For $\ell \geq 1$, let

$$\begin{aligned} \mathbf{a}(\ell) &= [a_1(\ell), \dots, a_\ell(\ell)]' = \arg \min_{(c_1, \dots, c_\ell)' \in \mathbb{R}^\ell} \mathbb{E}(\epsilon_t - c_1 \epsilon_{t-1} - \dots - c_\ell \epsilon_{t-\ell})^2, \\ \sigma^2(\ell) &= \mathbb{E}[\epsilon_t - a_1(\ell) \epsilon_{t-1} - \dots - a_\ell(\ell) \epsilon_{t-\ell}]^2, \end{aligned} \quad (2.5)$$

and $\sigma^2(0) = \mathbb{E}\epsilon_t^2$. Then the inverse matrix can be written as $\boldsymbol{\Sigma}^{-1} = \mathbf{F}' \mathbf{S}^{-1} \mathbf{F}$, where $\mathbf{S} = \text{diag}(\sigma^2(0), \sigma^2(1), \dots, \sigma^2(T-1))$, and $\mathbf{F} = (F_{ij}, 1 \leq i, j \leq T)$

with

$$F_{ij} = \begin{cases} 0, & \text{if } i < j, \\ 1, & \text{if } i = j, \\ -a_{i-j}(i-1), & \text{if } 2 \leq i \leq T, 1 \leq j \leq i-1. \end{cases} \quad (2.6)$$

If one assumes $\{\epsilon_t\}$ is weakly dependent, those elements that are far below the main diagonal of \mathbf{F} are nearly 0. The inverse matrix Σ^{-1} thus can be approximated by a truncated version that has less parameters. This method is previously adopted by Cheng et al. (2015), Ing et al. (2016) and Lin and Reuvers (2020b) with multiple applications as mentioned. The banded inverse autocovariance matrix (BIAM) is constructed as

$$\Sigma^{-1}(q) = \mathbf{F}'(q)\mathbf{S}^{-1}(q)\mathbf{F}(q), \quad (2.7)$$

where $q \ll T$ is called the banding parameter, $\mathbf{S}(q) = \text{diag}(\sigma^2(0), \sigma^2(1), \dots, \sigma^2(q), \dots, \sigma^2(q))$ and $\mathbf{F}(q) = (F_{ij}(q), 1 \leq i, j \leq T)$ with

$$F_{ij}(q) = \begin{cases} 0, & \text{if } i < j \text{ or } \{q+1 < i \leq T, 1 \leq j \leq i-q-1\}, \\ 1, & \text{if } i = j, \\ -a_{i-j}(i-1), & \text{if } 2 \leq i \leq q, 1 \leq j \leq i-1, \\ -a_{i-j}(q), & \text{if } q+1 \leq i \leq T, i-q \leq j \leq i-1. \end{cases} \quad (2.8)$$

Under suitable assumptions (given in the next section), the approximation error is negligible, i.e. $\|\Sigma^{-1}(q) - \Sigma^{-1}\| = o(1)$. To compare the MCD and the BIAM, an illustrative example is provided as follows.

Example 2.1

For $T = 4$, the MCD gives

$$\mathbf{F} = \begin{pmatrix} 1 & & & \\ -a_1(1) & 1 & & \\ -a_2(2) & -a_1(2) & 1 & \\ -a_3(3) & -a_2(3) & -a_1(3) & 1 \end{pmatrix} \quad \text{and} \quad \mathbf{S} = \begin{pmatrix} \sigma^2(0) & & & \\ & \sigma^2(1) & & \\ & & \sigma^2(2) & \\ & & & \sigma^2(3) \end{pmatrix}.$$

Consider $\mathbf{y}' \Sigma^{-1} \mathbf{y} = (\mathbf{F} \mathbf{y})' \mathbf{S}^{-1} (\mathbf{F} \mathbf{y})$. The premultiplications of \mathbf{y} by \mathbf{F} implies filtering, and \mathbf{S}^{-1} applies scaling. Moreover, suppose $q = 2$, the BIAM leads to

$$\mathbf{F}(q) = \begin{pmatrix} 1 & & & \\ -a_1(1) & 1 & & \\ -a_2(2) & -a_1(2) & 1 & \\ & -a_2(2) & -a_1(2) & 1 \end{pmatrix} \quad \text{and} \quad \mathbf{S}(q) = \begin{pmatrix} \sigma^2(0) & & & \\ & \sigma^2(1) & & \\ & & \sigma^2(2) & \\ & & & \sigma^2(2) \end{pmatrix}.$$

There is clearly a reduction in the number of parameters.

Without asymptotic loss, we will use $\Sigma^{-1}(q)$ to construct our infeasible GLS estimators. Let

$$\text{SSR}(k, \gamma; \Sigma^{-1}(q)) = (\mathbf{y} - \mathbf{X}_k \gamma)' \Sigma^{-1}(q) (\mathbf{y} - \mathbf{X}_k \gamma) \quad (2.9)$$

be the infeasible GLS sum of squared errors function. As seen from Example 2.1, the loss function (2.9) is essentially a weighted sum of filtered errors. The estimators of (k_0, γ_0) can be constructed by jointly minimizing the function (2.9). It is computationally less heavy through concentration. Given k , the infeasible GLS estimator of γ_0 is constructed as

$$\widehat{\gamma}(k; \Sigma^{-1}(q)) = [\mathbf{X}'_k \Sigma^{-1}(q) \mathbf{X}_k]^{-1} \mathbf{X}'_k \Sigma^{-1}(q) \mathbf{y}. \quad (2.10)$$

If it is not specifically mentioned, we use $\widehat{\gamma}_k = \widehat{\gamma}(k; \Sigma^{-1}(q))$. The infeasible GLS break date estimator is now given as

$$\begin{aligned} \hat{k} &= \arg \min_{k \in \Gamma_T} \text{SSR}(k; \widehat{\gamma}_k, \Sigma^{-1}(q)) \\ &= \arg \max_{k \in \Gamma_T} \left\{ \text{SSR}(k_0; \widehat{\gamma}_{k_0}, \Sigma^{-1}(q)) - \text{SSR}(k; \widehat{\gamma}_k, \Sigma^{-1}(q)) \right\}, \end{aligned} \quad (2.11)$$

where Γ_T is a subset of $\{1, \dots, T\}$. Finally, the coefficient estimates are obtained via the plug-in of \hat{k} , i.e. $\widehat{\gamma}_{\hat{k}}$. Two remarks are in place.

Remark 2.1

Our model above is also studied by Perron and Zhu (2005). Alternatively, one can also use a discontinuous trend function $t\mathbf{1}_{\{t>k\}}$ instead of $(t - k)\mathbf{1}_{\{t>k\}}$ to model trend breaks, see e.g., Bai (1997) and Bai et al. (1998). Note that

$$\mu + \beta t + \mu_b \mathbf{1}_{\{t>k\}} + \beta_b t \mathbf{1}_{\{t>k\}} = \mu + \beta t + (\mu_b + \beta_b k) \mathbf{1}_{\{t>k\}} + \beta_b (t - k) \mathbf{1}_{\{t>k\}}. \quad (2.12)$$

Thus, as mentioned in Perron and Zhu (2005), these two specifications yield the same estimation results of parameters k , μ , β and β_b in finite samples. However, the limiting approximations using either of the models can be rather diverse even under a common asymptotic framework; see the comparisons by Deng and Perron (2006). As seen from the right hand side in (2.12), the level shift $\mu_b + \beta_b k$ would be drastically larger than the slope change β_b in theory as T increases. Yet, the continuous trend specification is not restricted to this intrinsic property, and is theoretically more flexible in this sense. For this reason, in situations where level shifts are relatively small, it leads to a better approximation under suitable conditions.

Remark 2.2

We model a general Σ^{-1} directly on the contrary to the standard textbook strategy for GLS, namely estimating Σ first and then taking the inverse. This conventional approach, as discussed in Lin and Reuvers (2020b), often poses some difficulties in GLS. For instance, Greene (2012) estimated Σ by imposing certain special structures, which restricts the applications of GLS in many empirical situations. More generally, Wu and Pourahmadi (2009) and McMurry and Politis (2010) proposed estimators of Σ by banding or tapering sample autocovariance matrices. Unfortunately, their estimators cannot be guaranteed to be invertible in finite samples without specific treatments statistically.

2.3 Asymptotic Theory

The asymptotic results of infeasible GLS estimators are presented in this section. Our analysis is carried out under the following set of assumptions.

Assumption 2.1

Put $\Gamma_T = \{n_{1T}, n_{1T} + 1, \dots, T - n_{2T}\}$, where $n_{1T} = [\delta_1 T]$ and $n_{2T} = [\delta_2 T]$ for arbitrarily small numbers $\delta_1, \delta_2 \in (0, 1/2)$. We assume $k_0 = [\lambda_0 T]$ with $\lambda_0 \in (\delta_1, 1 - \delta_2)$.

Assumption 2.2

Let the break sizes $(\mu_{b,0}, \beta_{b,0}) = (d_{\mu,0} v_T, d_{\beta,0} v_T^3)$, where $d_{\mu,0}$ and $d_{\beta,0}$ are constants with $|d_{\mu,0}| + |d_{\beta,0}| > 0$. We assume $v_T > 0$ with $v_T \rightarrow 0$, $\frac{\sqrt{T} v_T}{\sqrt{\ln T}} \rightarrow \infty$ as $T \rightarrow \infty$.

Assumption 2.3

Suppose $A(z) \neq 0$ for all $|z| \leq 1$, where $A(\cdot)$ is defined in (2.3). The coefficients of $A(z)$ satisfy the summability condition $\sum_{j=0}^{\infty} j^2 |a_j| < \infty$. The process $\{\eta_t\}$ is a stationary ergodic martingale difference sequence (m.d.s.) w.r.t. its natural filtration $\mathcal{F}_t = \sigma(\eta_s, s \leq t)$, where $\mathbb{E}(\eta_t^2 | \mathcal{F}_{t-1}) = \sigma^2$ and $\sup_{-\infty < t < \infty} \mathbb{E}(\eta_t^4 | \mathcal{F}_{t-1}) \leq C$ with probability 1.

Assumption 2.4

The banding parameter $q \equiv q(T)$ satisfies $\frac{1}{q} + \frac{q}{T} \rightarrow 0$ as $T \rightarrow \infty$.

The trimming for Γ_T ensures the invertibility of $\mathbf{X}'_k \Sigma^{-1}(q) \mathbf{X}_k$. Assumption 2.1 also requires k_0 to be bounded away from the endpoints for specification purpose. It is a conventional assumption in the relevant literature, e.g. Bai (1997) and Bai and Perron (1998).

We assume the break sizes shrink to zero in Assumption 2.2, at the rates of v_T and v_T^3 . Clearly, they cannot be too small for identifiability. The same requirement for v_T is adopted in Bai (1994). These rates provide appropriate scalings so that a functional central limit theorem (FCLT) can be later applied. The signals from both level and trend breaks can be captured under these rates in our limiting distribution unlike the asymptotics of Deng and Perron (2006). This shrinking framework can eliminate the dependency of asymptotic

distributions on the exact distributions of disturbances, and make a simple limit approximation to be possible. As it is permitted to have only one break in either level or slope, the DGP is therefore nested in our model in these cases.

Assumption 2.3 implies the spectral density function of $\{\epsilon_t\}$ is bounded and bounded away from zero. Using the inversion formula, see e.g. Proposition 4.5.3 of Brockwell and Davis (1991) or page 6 in Jentsch and Politis (2015), it is true that $\|\Sigma\| < \infty$ and $\|\Sigma^{-1}\| < \infty$. This assumption also enables us to apply the FCLT in Theorem 3.15 of Phillips and Solo (1992), as well as the First Moment Bound Theorem of Findley and Wei (1993) to obtain sharp bounds for the moments of ϵ_t in quadratic forms. The purpose of Assumption 2.4 is twofold. First, we assume the banding parameter q grows with sample size for reducing the loss incurred from truncations. Second, we require the number of parameters in $\Sigma^{-1}(q)$ not to increase too fast. Under this condition, any q rows of vectors or matrices such as $F(q)\epsilon$ and $F(q)X_k$ are usually negligible with appropriate scalings, which provides some convenience in the proofs. Finally, Assumptions 3 and 4 together are sufficient to imply $\|\Sigma^{-1}(q) - \Sigma^{-1}\| = o(1)$.

2.3.1 Asymptotics of Infeasible GLS Break Date Estimator

We first establish the consistency and the rate of convergence for the break date estimator.

Proposition 1

Under Assumptions 2.1 to 2.4, $v_T^2(\hat{k} - k_0) = O_p(1)$.

Proposition 1 implies that the estimated break fraction $\hat{\lambda} = \hat{k}/T$ is Tv_T^2 -consistent. Intuitively speaking, if v_T decays to zero faster, one would have less signals for identifying break dates, leading to less accurate estimates. The same result also holds in various settings such as Proposition 1 of Bai (1997) and Lemma 1 of Qu and Perron (2007). Proposition 1 allows us to analyze

the limiting distribution in a neighborhood of k_0 with length $M_0 v_T^{-2}$ for some $M_0 > 0$. Since the neighborhood increases as T grows, a FCLT is possible to be applied to eliminate the reliance on error distributions.

Theorem 2.1

Under Assumptions 2.1 to 2.4,

$$\begin{aligned} \frac{\beta_{b,0}^2}{3\sigma_\epsilon^2}(\hat{k} - k_0)^3 + \frac{\mu_{b,0}\beta_{b,0}}{\sigma_\epsilon^2}(\hat{k} - k_0)^2 + \frac{\mu_{b,0}^2}{\sigma_\epsilon^2}(\hat{k} - k_0) \\ \xrightarrow{\mathcal{D}} \arg \max_{v \in (-\infty, \infty)} \left\{ W(v) - \frac{|v|}{2} \right\} =: \mathcal{W} \end{aligned} \quad (2.13)$$

where $\sigma_\epsilon^2 = \sigma^2 A(1)^{-2}$ is the long-run variance of ϵ_t , $W(v) = W_1(v)$ for $v \geq 0$ and $W(v) = W_2(-v)$ for $v < 0$, and $W_i(s)$, $i = 1, 2$, are independent standard Brownian motions defined on $[0, \infty)$.

Theorem 2.1 has several implications. First of all, the left hand side of (2.13) is a cubic polynomial transformation of $\hat{k} - k_0$, which is rather different from the ‘standard’ form in the literature. It is typically found, for example, Proposition 3 in Bai (1997) and Theorem 4 in Bai et al. (1998), that certain *linear* functions of $\hat{k} - k_0$ converge to \mathcal{W} . As will be seen later, this new-found transformation captures the asymmetric and bimodal features in finite samples. Second, there are two special cases given in the following corollary.

Corollary 2.1

The limiting distribution (2.13) implies

$$\frac{\mu_{b,0}^2}{\sigma_\epsilon^2}(\hat{k} - k_0) \xrightarrow{\mathcal{D}} \mathcal{W}, \quad \text{if} \quad d_{\beta,0} = 0, \quad (2.14)$$

$$\frac{\beta_{b,0}^2}{3\sigma_\epsilon^2}(\hat{k} - k_0)^3 \xrightarrow{\mathcal{D}} \mathcal{W}, \quad \text{if} \quad d_{\mu,0} = 0. \quad (2.15)$$

Suppose there is no trend break in the DGP ($d_{\beta,0} = 0$), nevertheless, it is

allowed in the model specification (2.1). The asymptotic distribution (2.14) coincides with Equation (14) of Bai (1997) for a change-in-mean model. In other words, there is no efficiency gain by ‘knowing’ no trend break. It is strikingly different from the case $d_{\mu,0} = 0$. If the model excludes a level break, the limiting approximation is often unimodal, see Item 2 in Theorem 4 of Perron and Zhu (2005) or Theorem 3(1) of Deng and Perron (2006), irrespective of fixed or shrinking break sizes are assumed. Instead, the distribution (2.15) is bimodal, see Figure 2.1 (a) below. An intuitive reason is as follows. When a level break is included, around k_0 , any fluctuations in observations caused by randomness can be recognized as a break in the intercept. Consequently, the probability of incorrectly detecting a break before or after k_0 is higher for the model that excludes a level change. In this case, ‘knowing’ no level break brings asymptotic gain.

It is not hard to obtain the limiting probability density function (pdf) of $\hat{k} - k_0$. The properties of \mathcal{W} are well-studied, see, e.g., Bhattacharya and Brockwell (1976) and Picard (1985). It has the pdf f :

$$f(x) = \frac{3}{2}e^{|x|}\Phi\left(-\frac{3\sqrt{|x|}}{2}\right) - \frac{1}{2}\Phi\left(-\frac{\sqrt{|x|}}{2}\right), \quad x \in \mathbb{R}, \quad (2.16)$$

where Φ is the cumulative standard normal distribution function. The asymptotic pdf of $\hat{k} - k_0$ now follows from Theorem 2.1.

Corollary 2.2

Denote by \mathbb{T} a cubic polynomial transformation:

$$\mathbb{T}(x; \mu_{b,0}, \beta_{b,0}, \sigma_\epsilon^2) = \frac{1}{\sigma_\epsilon^2} \left[\frac{1}{3} \beta_{b,0}^2 x^3 + \mu_{b,0} \beta_{b,0} x^2 + \mu_{b,0}^2 x \right]. \quad (2.17)$$

If there is no risk of confusion, we write $\mathbb{T}(x) = \mathbb{T}(x; \mu_{b,0}, \beta_{b,0}, \sigma_\epsilon^2)$ and $\mathbb{T}'(x) = \sigma_\epsilon^{-2} (\beta_{b,0} x + \mu_{b,0})^2$. Under the conditions of Theorem 2.1, the asymptotic pdf of $\hat{k} - k_0$ is

$$k(x; \mu_{b,0}, \beta_{b,0}, \sigma_\epsilon^2) = f[\mathbb{T}(x)] \mathbb{T}'(x), \quad x \in \mathbb{R}. \quad (2.18)$$

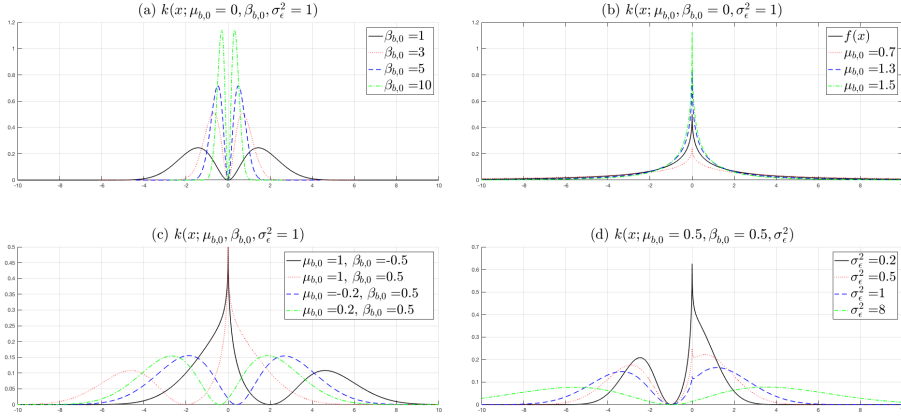


Figure 2.1: The asymptotic pdf's f and $k(\cdot; \mu_{b,0}, \beta_{b,0}, \sigma_\epsilon^2)$.

We plot the densities f and $k(\cdot; \mu_{b,0}, \beta_{b,0}, \sigma_\epsilon^2)$ with varying parameters in Figure 2.1. Note that, since f is positive, $k(x; \mu_{b,0}, \beta_{b,0}, \sigma_\epsilon^2) = 0$ only at $x = -\mu_{b,0}/\beta_{b,0}$ if $d_{\beta,0} \neq 0$. From Figure 2.1 (a), it is noticeable how the cubic term captures the bimodality as widely reported in the literature, for instance, Figure 6 in Perron and Zhu (2005), Figure 1 in Deng and Perron (2006), and Figure 1 in Harvey and Leybourne (2014). Figure 2.1 (b) considers the cases of $d_{\beta,0} = 0$. The asymptotic pdf of $\hat{k} - k_0$ preserves the features of f , that is, unimodality, symmetry, continuity and non-differentiability at $x = 0$. When $d_{\mu,0}d_{\beta,0} \neq 0$, as shown in Figure 2.1 (c), the asymptotic pdf carries on the shapes together from the previous two cases. Specially, it is no longer symmetric when both level and trend breaks present, which is also found by the authors mentioned above. Figure 2.1 (d) shows that the influences of break signal-to-noise ratios from level and trend slope are different. Due to the fact that level break is generally harder to be discriminated from randomness, when errors are less noisy, the rise in detection probability mainly stems from level break.

2.3.2 Asymptotic Equivalence of LS Break Date Estimators

There are many papers reporting the finite sample and asymptotic equivalence among LS estimators under various circumstances in the literature, for instance,

see Grenander and Rosenblatt (1957) and Canjels and Watson (1997) for trending models, Park and Phillips (1988) and Lin and Reuvers (2020b) for cointegrating regressions. In this section, we give a brief comparison of the proposed infeasible GLS estimator \hat{k} and the OLS and QD estimators that are developed by Perron and Zhu (2005) and Harvey and Leybourne (2014) respectively. More specifically, define

$$\Delta_\rho = \mathbf{I}_T - \rho \mathbf{L}_T, \quad |\rho| < 1, \quad (2.19)$$

where \mathbf{I}_T is the T -dimensional identity matrix, and $\mathbf{L}_T = (L_{ij})$ is a lower shift matrix such that $L_{ij} = 0$ if $i \neq j+1$, and $L_{ij} = 1$ if $i = j+1$. The OLS estimator is based on $\text{SSR}(k; \widehat{\gamma}_{k,OLS}, \mathbf{I}_T)$ with $\widehat{\gamma}_{k,OLS} := \widehat{\gamma}(k; \mathbf{I}_T)$. Similarly, the QD is obtained using $\text{SSR}(k; \widehat{\gamma}_{k,QD}, \Delta'_\rho \Delta_\rho)$, where $\widehat{\gamma}_{k,QD} := \widehat{\gamma}(k; \Delta'_\rho \Delta_\rho)$.

We make two observations. First, the GLS estimator \hat{k} can be equal to the OLS and QD estimators in finite samples. For example, \hat{k} is equivalent to the QD estimator if $\{\epsilon_t - \rho\epsilon_{t-1}\}$ is serially uncorrelated. In addition, we obtain the OLS estimator for $\rho = 0$. Second, $\Sigma^{-1}(q)$ vanishes into σ_ϵ^{-2} in the asymptotic distribution (2.13), suggesting that the GLS, QD and OLS break date estimators might be asymptotically equivalent. In fact, it can be proved that the criterion function of GLS in (2.11) only differs from the counterparts of QD and OLS by a multiplicative constant asymptotically. As a result, we obtain the asymptotic equivalence among these LS estimators.

Proposition 2

Under Assumptions 2.1 to 2.4, OLS, QD and GLS break date estimators share the same limiting distribution (2.13).

2.3.3 Asymptotics of Infeasible GLS Coefficient Parameters

Next to inference on the break date, our approach also allows to conduct inference on the coefficients. Here we provide the limit distributions of the estimators.

Theorem 2.2

Write $\widehat{\gamma}_{\hat{k}} = (\hat{\mu}_{\hat{k}}, \hat{\beta}_{\hat{k}}, \hat{\mu}_{b,\hat{k}}, \hat{\beta}_{b,\hat{k}})'$ explicitly. Let $D_T = \text{diag}(T, T^3, T, T^3)$, and

$$\Omega_{\lambda}^{-1} = \begin{pmatrix} \frac{4}{\lambda} & -\frac{6}{\lambda^2} & \frac{2}{\lambda} & \frac{6}{\lambda^2} \\ -\frac{6}{\lambda^2} & \frac{12}{\lambda^3} & -\frac{6}{\lambda^2} & -\frac{12}{\lambda^3} \\ \frac{2}{\lambda} & -\frac{6}{\lambda^2} & \frac{4}{\lambda(1-\lambda)} & \frac{6(1-2\lambda)}{\lambda^2(1-\lambda)^2} \\ \frac{6}{\lambda^2} & -\frac{12}{\lambda^3} & \frac{6(1-2\lambda)}{\lambda^2(1-\lambda)^2} & \frac{12(3\lambda^2-3\lambda+1)}{\lambda^3(1-\lambda)^3} \end{pmatrix}. \quad (2.20)$$

Suppose Assumptions 2.1 to 2.4 hold.

(i) In general, $D_T^{1/2}(\widehat{\gamma}_{k_0} - \gamma_0) \xrightarrow{\mathcal{D}} \mathcal{N}(\mathbf{0}, \sigma_{\epsilon}^2 \Omega_{\lambda_0}^{-1})$ and

$$D_T^{1/2}(\widehat{\gamma}_{\hat{k}} - \gamma_0) - \sqrt{T}\beta_{b,0}(\hat{k} - k_0) \begin{pmatrix} 0 \\ 0 \\ 1 \\ 0 \end{pmatrix} \xrightarrow{\mathcal{D}} \mathcal{N}(\mathbf{0}, \sigma_{\epsilon}^2 \Omega_{\lambda_0}^{-1}). \quad (2.21)$$

(ii) If $d_{\beta,0} = 0$, $D_T^{1/2}(\widehat{\gamma}_{\hat{k}} - \gamma_0)$ and $D_T^{1/2}(\widehat{\gamma}_{k_0} - \gamma_0)$ both converge in distribution to $\mathcal{N}(\mathbf{0}, \sigma_{\epsilon}^2 \Omega_{\lambda_0}^{-1})$. Moreover,

$$\sqrt{T}(\hat{\mu}_{b,\hat{k}} - \mu_{b,0}) \xrightarrow{\mathcal{D}} \mathcal{N}\left(0, \frac{4}{\lambda_0(1-\lambda_0)}\right). \quad (2.22)$$

(iii) If $d_{\beta,0} \neq 0$,

$$v_T^{-1}(\hat{\mu}_{b,\hat{k}} - \mu_{b,0}) = d_{\beta,0} v_T^2(\hat{k} - k_0) + o_p(1) \xrightarrow{\mathcal{D}} d_{\beta,0} \arg \max_{s \in \mathbb{R}} G(s), \quad (2.23)$$

where

$$G(s) = \begin{cases} \sigma_{\epsilon}^{-2} \left[\frac{1}{3} d_{\beta,0}^2 s^3 + d_{\mu,0} d_{\beta,0} s^2 + d_{\mu,0}^2 s \right] - 2\sigma_{\epsilon}^{-1} \left[d_{\mu,0} W_2(-s) + d_{\beta,0} \int_0^{-s} r dW_2(r) \right], & s < 0, \\ -\sigma_{\epsilon}^{-2} \left[\frac{1}{3} d_{\beta,0}^2 s^3 + d_{\mu,0} d_{\beta,0} s^2 + d_{\mu,0}^2 s \right] - 2\sigma_{\epsilon}^{-1} \left[d_{\mu,0} W_1(s) + d_{\beta,0} \int_0^s r dW_1(r) \right], & s \geq 0, \end{cases}$$

and $W_1(s)$ and $W_2(s)$ are independent standard Brownian motions defined on $[0, \infty)$.

Some additional remarks are in order. First, as argued for \hat{k} in Proposition 2, OLS, QD and GLS estimators of coefficient parameters are asymptotically equivalent, which can be confirmed by comparing our Theorem 2.2 with part 4 in Theorem 6 of Perron and Zhu (2005). Similar conclusions can be also found in Grenander and Rosenblatt (1957) and Canjels and Watson (1997) in trending models without breaks. Second, $\hat{\mu}_{b,\hat{k}}$ has rather different asymptotic behaviors, depending on the value of $d_{\beta,0}$. If $d_{\beta,0} = 0$, $\hat{\mu}_{b,\hat{k}}$ is \sqrt{T} -consistent. In this case, $\widehat{\gamma}_{\hat{k}}$ has the same limiting distributions as if the true break date were known. Thus, the estimation of the coefficient parameters should not be sensitive to the precision of break date estimates. This is in line with existing results for change-in-mean models. It is also interesting to see that $\hat{\mu}_{b,\hat{k}}$ and $\hat{\beta}_{b,\hat{k}}$ are asymptotically independent at $\lambda_0 = 1/2$ because $\frac{6(1-2\lambda_0)}{\lambda_0^2(1-\lambda_0)^2} = 0$. If $d_{\beta,0} \neq 0$, $\hat{\mu}_{b,\hat{k}} - \mu_{b,0}$ vanishes at rate v_T^{-1} in light of (2.23), which is slower than \sqrt{T} , and is affected by the performance of \hat{k} through the factor $d_{\beta,0}$. As discussed below Corollary 2.1, the inclusion of level break in the model may likewise has an influence on the precision of \hat{k} . This is called the ‘feedback’ effect by Perron and Zhu (2005), see their Section 5 for detailed discussions. Third, (2.21) provides separate inference on $\beta_{b,0}$. Moreover, one could conduct joint tests on nonzero restrictions on $\mu_{b,0}$ and zero restrictions on $\beta_{b,0}$. Due to the contaminating term $\hat{k} - k_0$ in (2.21), it is unclear how to conduct joint tests for nonzero restrictions on both $\mu_{b,0}$ and $\beta_{b,0}$. Finally, our large sample theory relies on Assumption 2.2 which assumes the break sizes of level and trend slope cannot be simultaneously zero. In view of this, it is not plausible to test zero restrictions on both $\mu_{b,0}$ and $\beta_{b,0}$ within our framework.

2.3.4 Estimation of the Banded Inverse Autocovariance Matrix

Up to now, we have assumed $\Sigma^{-1}(q)$ is known. In this section, we discuss the estimation of the BIAM, and show the consistency results. Since the errors

are unobservable, we can only estimate $\Sigma^{-1}(q)$ via regression residuals. To obtain residuals, we can first plug in the OLS or QD break date estimates discussed in Section 2.3.2. It is necessary to select the QD parameter ρ for implementing the latter approach. Following Harvey and Leybourne (2014), a value of ρ can be conveniently chosen together with the break date estimate by jointly minimizing sum of squared residuals across Γ_T and a grid of the interval $(-1, 1)$. Accordingly, let

$$\mathbf{r}_m = \{ \bar{\rho}_i : -1 < \bar{\rho}_1 < \bar{\rho}_2 < \cdots < \bar{\rho}_m < 1 \} \quad (2.24)$$

be a grid of $(-1, 1)$. Recall $\Delta_\rho = \mathbf{I}_T - \rho \mathbf{L}_T$, where \mathbf{L}_T is a lower shift matrix. A two-step procedure can be executed as follows.

(S1) Take

$$(\hat{k}_{r_m, T}, \hat{\rho}_{r_m, T}) = \arg \min_{(k, \bar{\rho}) \in \Gamma_T \times \mathbf{r}_m} \text{SSR} \left\{ k; \widehat{\gamma} \left(k; \Delta'_{\bar{\rho}} \Delta_{\bar{\rho}} \right), \Delta'_{\bar{\rho}} \Delta_{\bar{\rho}} \right\}, \quad (2.25)$$

and define $(\hat{k}_r, \hat{\rho}_r) = (\hat{k}_{r_m, T}, \hat{\rho}_{r_m, T})$ by suppressing the dependence on m and T .

(S2) Conditional on \hat{k}_r , construct $\widehat{\gamma}_{\hat{k}_r} = \left(\mathbf{X}'_{\hat{k}_r} \mathbf{X}_{\hat{k}_r} \right)^{-1} \mathbf{X}'_{\hat{k}_r} \mathbf{y}$ and compute the residuals,

$$\widehat{\varepsilon}(\hat{k}_r) = [\hat{\varepsilon}_1(\hat{k}_r), \cdots, \hat{\varepsilon}_T(\hat{k}_r)]' = \mathbf{y} - \mathbf{X}_{\hat{k}_r} \widehat{\gamma}_{\hat{k}_r}. \quad (2.26)$$

For $1 \leq \ell \leq q$, a sample counterpart of (2.5) is given as

$$\begin{aligned} \widehat{\mathbf{a}}(\ell, \hat{k}_r) &= [\hat{a}_1(\ell, \hat{k}_r), \cdots, \hat{a}_\ell(\ell, \hat{k}_r)]' \\ &= \arg \min_{(c_1, \cdots, c_\ell)' \in \mathbb{R}^\ell} \sum_{t=\ell+1}^T [\hat{\varepsilon}_t(\hat{k}_r) - c_1 \hat{\varepsilon}_{t-1}(\hat{k}_r) - \cdots - c_\ell \hat{\varepsilon}_{t-\ell}(\hat{k}_r)]^2 \\ \widehat{\sigma}^2(\ell, \hat{k}_r) &= \frac{1}{T-\ell} \sum_{t=\ell+1}^T [\hat{\varepsilon}_t(\hat{k}_r) - \hat{a}_1(\ell, \hat{k}_r) \hat{\varepsilon}_{t-1}(\hat{k}_r) - \cdots - \hat{a}_\ell(\ell, \hat{k}_r) \hat{\varepsilon}_{t-\ell}(\hat{k}_r)]^2 \end{aligned} \quad (2.27)$$

and $\widehat{\sigma}^2(0, \hat{k}_r) = \frac{1}{T} \sum_{t=1}^T \hat{\epsilon}_t^2(\hat{k}_r)$. Plug these estimates into $F(q)$ and $S(q)$ respectively,

$$\begin{aligned}\widehat{F}(q, \hat{k}_r) &= (\widehat{F}_{ij}(q, \hat{k}_r), 1 \leq i, j \leq T) \\ \widehat{S}(q, \hat{k}_r) &= \text{diag}(\widehat{\sigma}^2(0, \hat{k}_r), \widehat{\sigma}^2(1, \hat{k}_r), \dots, \widehat{\sigma}^2(q, \hat{k}_r), \dots, \widehat{\sigma}^2(q, \hat{k}_r)).\end{aligned}$$

The resulting BIAM estimator is

$$\widehat{\Sigma}^{-1}(q, \hat{k}_r) = \widehat{F}'(q, \hat{k}_r) \widehat{S}^{-1}(q, \hat{k}_r) \widehat{F}(q, \hat{k}_r). \quad (2.28)$$

Note that (2.25) implies essentially the OLS break date estimator if $r_m = \{0\}$. For $m > 1$, (2.25) leads to the QD estimator. To derive the consistency of the BIAM estimator, stricter assumptions on q are required.

Assumption 2.5

Assume $q = q(T)$ with $\frac{1}{q} + \frac{q^3}{T} \mathbf{1}_{\{d_{\beta,0}=0\}} + q^3 v_T^2 \mathbf{1}_{\{d_{\beta,0} \neq 0\}} \rightarrow 0$ as $T \rightarrow \infty$.

When $d_{\beta,0} \neq 0$, as $T \rightarrow \infty$, the third term $q^3 v_T^2 \rightarrow 0$ in Assumption 2.5 implies $q^3/T \rightarrow 0$ using $\sqrt{T}v_T \rightarrow \infty$ (Assumption 2.2), and further implies Assumption 2.4, and thus is the strongest. In this case, we could drop the second term in the assumption when $d_{\beta,0}$ is known to be nonzero, i.e. $q^{-1} + q^3 v_T^2 \xrightarrow{T \rightarrow \infty} 0$. When $d_{\beta,0} = 0$ (with change-in-mean only), the assumption reduces to a common form $q^{-1} + q^3/T \xrightarrow{T \rightarrow \infty} 0$ (Berk (1974), Lin and Reuvers (2020b)). Therefore, the BIAM estimator is not affected by the first-stage break date estimates asymptotically in this scenario. The following proposition gives the consistency of estimated BIAM, which will be used to show that Feasible GLS (FGLS) and infeasible GLS estimators share the same limiting distributions.

Theorem 2.3

If Assumptions 2.1 to 2.3 and 2.5 hold, then $\|\widehat{\Sigma}^{-1}(q, \hat{k}_r) - \Sigma^{-1}\| \xrightarrow{\mathcal{P}} 0$.

Remark 2.3

By the asymptotic equivalence, under Assumptions 2.1 to 2.4, it follows that $v_T^2(\hat{k}_r - k_0) = O_p(1)$, and $\mathbf{D}_T^{1/2}(\widehat{\gamma}_{\hat{k}_r} - \gamma_0) = O_p(1)$ provided $d_{\beta,0} = 0$, and similarly, $\mathbf{D}_{v_T, T}^{1/2}(\widehat{\gamma}_{\hat{k}_r} - \gamma_0) = O_p(1)$ if $d_{\beta,0} \neq 0$, where $\mathbf{D}_{v_T, T} = \text{diag}(T, T^3, v_T^{-2}, T^3)$.

2.3.5 Asymptotics of Feasible GLS Parameter Estimators

FGLS estimators of parameters can readily be obtained by the plug-in of the estimated BIAM. Similarly as (2.10), we construct

$$\widehat{\gamma}(k; \widehat{\Sigma}^{-1}(q, \hat{k}_r)) = [\mathbf{X}'_k \widehat{\Sigma}^{-1}(q, \hat{k}_r) \mathbf{X}_k]^{-1} \mathbf{X}'_k \widehat{\Sigma}^{-1}(q, \hat{k}_r) \mathbf{y}. \quad (2.29)$$

We denote $\widehat{\gamma}_{k,f} = \widehat{\gamma}(k; \widehat{\Sigma}^{-1}(q, \hat{k}_r))$, where the subscript f stands for ‘feasible’ to discriminate from $\widehat{\gamma}_k$ defined in (2.10). Then the feasible GLS break date estimator can be given as

$$\hat{k}_f = \arg \min_{k \in \Gamma_T} \text{SSR}(k; \widehat{\gamma}_{k,f}, \widehat{\Sigma}^{-1}(q, \hat{k}_r)), \quad (2.30)$$

and finally the FGLS estimator of coefficient parameters is $\widehat{\gamma}_{\hat{k}_f, f}$. The next proposition illustrates that the asymptotic results of infeasible GLS are also valid for FGLS, and can be used for inference.

Proposition 3

Under Assumptions 2.1 to 2.3 and 2.5, the results in Propositions 1-2 and Theorems 2.1-2.2 hold for FGLS estimators \hat{k}_f and $\widehat{\gamma}_{\hat{k}_f, f} = (\hat{\mu}_{\hat{k}_f}, \hat{\beta}_{\hat{k}_f}, \hat{\mu}_{b, \hat{k}_f}, \hat{\beta}_{b, \hat{k}_f})'$ correspondingly.

2.4 Confidence Sets for Break Dates

It is common to construct confidence intervals for break dates via the inversion of Wald-type (or t) statistics, see, e.g., Section D of Bai (1997). Simulation studies often report poor coverage of the Wald-type interval when there are

small breaks, see e.g. Chang and Perron (2018). Another option to form confidence sets is by inverting multiple locally best invariant (LBI) tests, see Elliott and Müller (2007) in stationary models, and Harvey and Leybourne (2015) in trending models. It is widely recorded that the LBI-based confidence sets yield accurate or conservative coverage with small break sizes. However, they can be too lengthy to be informative, for instance, see Eo and Morley (2015), Chang and Perron (2018) and Yamamoto (2018). Furthermore, based on a local asymptotic framework, the LBI tests also suffer from the nonmonotonic power problem that is well-known in the literature of structural change tests, i.e. the power of tests can decrease as the break sizes increase. As a result, the length of confidence sets can also increase as the break sizes increase. As suggested in the conclusion of Chang and Perron (2018), “In the context of structural change tests, one should be skeptical of the use of some local asymptotic frameworks...” These concerns motivate us to seek for a method that has robust performance in terms of both coverage and length. Moreover, the length shall improve given a larger break. In this section, we first discuss the construction of likelihood ratios (LR) based confidence sets and intervals by the inversion of multiple LR tests. As a benchmark, we also provide the Wald-type intervals using the asymptotic distribution in Theorem 2.1. Throughout, we denote q_α for a generic α_{th} quantile of certain distributions.

2.4.1 LR-Based Confidence Sets and Intervals

We consider a sequence of tests for $H_0 : k = k_0$ against $H_1 : k \neq k_0$, where $k \in \Gamma_T$. The (quasi-)log-LR tests statistic based on FGLS is constructed as

$$\text{LR}_T(k) = \text{SSR}(k; \widehat{\gamma}_{k,f}, \widehat{\Sigma}^{-1}(q, \hat{k}_r)) - \text{SSR}(\hat{k}_f; \widehat{\gamma}_{\hat{k}_f,f}, \widehat{\Sigma}^{-1}(q, \hat{k}_r)). \quad (2.31)$$

Theorem 2.4

Suppose Assumptions 2.1 to 2.3 and 2.5 hold. Under the null hypothesis,

$$\text{LR}_T(k_0) \xrightarrow{\mathcal{D}} \mathcal{L}, \quad (2.32)$$

where $\mathcal{L} = \max_{v \in (-\infty, \infty)} \{2W(v) - |v|\}$ with the distribution function $\mathbb{P}(\mathcal{L} \leq x) = (1 - e^{-x/2})^2$.

The limiting distribution \mathcal{L} is free of nuisance parameters such as the long-run variance σ_ϵ^2 , which is different from Hansen (2000) and Eo and Morley (2015). The serial correlation of error processes is automatically removed by the BIAM estimator, and is thus scaled to unit variance. The distribution of \mathcal{L} is studied by Bhattacharya and Brockwell (1976). The analytic $(1 - \alpha)_{th}$ quantile of \mathcal{L} can be easily computed as $q_{1-\alpha} = -2 \ln(1 - \sqrt{1 - \alpha})$. Moreover, let

$$C_{lr} = \{k : \text{LR}_T(k) \leq q_{1-\alpha}\} \quad (2.33)$$

be the LR-based confidence sets. Under the assumptions in Theorem 2.4, C_{lr} is asymptotically valid for a level α , that is $\mathbb{P}(k_0 \in C_{lr}) \rightarrow 1 - \alpha$. The sample paths $k \mapsto \text{LR}_T(k)$, $k \in \Gamma_T$, are often multimodal in finite samples. Thus, C_{lr} is possibly a disconnected confidence set. A reasonable modification into intervals suggested by Siegmund (1988) is

$$C_{lr}^* = [\min C_{lr}, \max C_{lr}]. \quad (2.34)$$

There are a few words on the implementation of constructing the sets/intervals above. One could first save the sequence $\{\text{SSR}(k; \widehat{\gamma}_{k,f}, \widehat{\Sigma}^{-1}(q, \hat{k}_r)), k \in \Gamma_T\}$ in the FGLS estimation of break date in (2.30) and then subtract it by its minimum. Since no nuisance parameters are present in \mathcal{L} , the sets/intervals can be immediately obtained by comparing the sequence of values with $q_{1-\alpha}$.

2.4.2 Wald-type Confidence Intervals

As a benchmark, we provide a Wald-type construction of confidence intervals using the asymptotics in Theorem 2.1. First, we can compute the inverse

function \mathbb{T}^{-1} of \mathbb{T} such that

$$\mathbb{T}^{-1}\left(x; \mu_{b,0}, \beta_{b,0}, \sigma_{\epsilon}^2\right) = \begin{cases} \frac{\left[3\sigma_{\epsilon}^2\beta_{b,0}x + \mu_{b,0}^3\right]^{1/3} - \mu_{b,0}}{\beta_{b,0}}, & \beta_{b,0} \neq 0, \\ \frac{\sigma_{\epsilon}^2}{\mu_{b,0}^2}x, & \beta_{b,0} = 0. \end{cases} \quad (2.35)$$

Second, substitute the corresponding FGLS estimates for the nuisance parameters, and denote by

$$\widehat{\mathbb{T}}^{-1}(x) = \mathbb{T}^{-1}\left(x; \hat{\mu}_{b,\hat{k}_f}, \hat{\beta}_{b,\hat{k}_f}, \widehat{\sigma}_{\epsilon,\hat{k}_f}^2\right),$$

where $\widehat{\sigma}_{\epsilon,\hat{k}_f}^2$ is a consistent estimator of the long-run variance σ_{ϵ}^2 , see, e.g., Berk (1974) or Andrews (1991). Third, let $c_{\alpha/2} < 0$ and $c_{1-\alpha/2} > 0$ be the quantiles of \mathcal{W} such that $\mathbb{P}(c_{\alpha/2} \leq \mathcal{W} \leq c_{1-\alpha/2}) = 1 - \alpha$.¹ Then a feasible $100(1 - \alpha)\%$ confidence interval is given as

$$C_{wald} = \left[\hat{k}_f - \left\lceil \widehat{\mathbb{T}}^{-1}(c_{1-\alpha/2}) \right\rceil - 1, \hat{k}_f - \left\lfloor \widehat{\mathbb{T}}^{-1}(c_{\alpha/2}) \right\rfloor + 1\right] \cap \Gamma_T, \quad (2.36)$$

where $\lceil x \rceil$ takes the least integer that is greater or equal to x , and $\lfloor x \rfloor$ denoting the greatest integer that is less than or equal to x . Taking the intersection with Γ_T is to avoid the intervals to include dates that are out of samples. In practice, whether $\beta_{b,0}$ is equal to zero is unknown. One could use the asymptotics in Theorem 2.2 to do inference for $\beta_{b,0}$ first and then determine which case in (2.35) to be applied.² Since the inference of slope coefficients is not our main consideration, we shall assume the information of $\beta_{b,0}$ is known in the simulations.

2.5 Prewhitening to Improve Finite-Sample Properties

The estimation and inference for break dates can be improved by employing a prewhitening procedure when there is considerable temporal dependence.

¹By the symmetry of \mathcal{W} , we have $c_{\alpha/2} = -c_{1-\alpha/2}$. For $\alpha = 1\%, 5\%, 10\%$, $c_{1-\alpha/2} = 19.762, 11.040, 7.690$, respectively.

²One may need to be aware of the issues resulting from conditional inference.

It essentially extracts the main serial dependence of $\{\epsilon_t\}$ out in the first place. Intuitively speaking, under a strong serial correlation, a good BIAM estimator requires a large banding parameter regardless of the sample size. Nevertheless, q is chosen to be small when T is small. The BIAM estimator is thus sensitive to the selection of q in this case. More specifically, instead of estimating Σ^{-1} , we estimate the inverse of

$$\Sigma_\rho = \mathbb{E} \left[(\Delta_\rho \epsilon)(\Delta_\rho \epsilon)' \right] \quad (2.37)$$

for some QD parameter $|\rho| < 1$. The QD errors $\Delta_\rho \epsilon = (\epsilon_t - \rho \epsilon_{t-1}, 1 < t \leq T)$ can be viewed as prewhitening ϵ_t by an AR(1) process. A similar idea can be also found in Andrews and Monahan (1992), in which a VAR-prewhitened procedure is used to construct the heteroskedasticity and autocorrelation (HAC) estimator developed in Andrews (1991), resulting in better finite sample results of t tests. Similarly, by substituting $\epsilon_t - \rho \epsilon_{t-1}$ for ϵ_t in Equations (2.5) to (2.8), we get the BIAM for Σ_ρ given as

$$\Sigma_\rho^{-1} = F'_\rho(q) S_\rho^{-1}(q) F_\rho(q). \quad (2.38)$$

We then modify the two-step procedure in Section 2.3.4 as follows.

(S1) The first step remains unchanged. Now define $\hat{\pi}_r = (\hat{k}_r, \hat{\rho}_r)$.

(S2) Conditional on $\hat{\pi}_r$, construct

$$\widehat{\gamma}_{\hat{\pi}_r} = \left[(\Delta_{\hat{\rho}_r} \mathbf{X}_{\hat{k}_r})' (\Delta_{\hat{\rho}_r} \mathbf{X}_{\hat{k}_r}) \right]^{-1} (\Delta_{\hat{\rho}_r} \mathbf{X}_{\hat{k}_r})' (\Delta_{\hat{\rho}_r} \mathbf{y})$$

and compute the QD residuals, w.r.t. $|\rho| < 1$,

$$\widehat{\epsilon}(\rho, \hat{\pi}_r) = [\hat{\epsilon}_1(\rho, \hat{\pi}_r), \dots, \hat{\epsilon}_T(\rho, \hat{\pi}_r)]' = \Delta_\rho [\mathbf{y} - \mathbf{X}_{\hat{k}_r} \widehat{\gamma}_{\hat{\pi}_r}]. \quad (2.39)$$

For $1 \leq \ell \leq q_T$, a sample counterpart of (2.5) is now given as

$$\begin{aligned}\widehat{\mathbf{a}}_\rho(\ell, \hat{\boldsymbol{\pi}}_r) &= [\hat{a}_{\rho,1}(\ell, \hat{\boldsymbol{\pi}}_r), \dots, \hat{a}_{\rho,\ell}(\ell, \hat{\boldsymbol{\pi}}_r)]' \\ &= \arg \min_{(c_1, \dots, c_\ell)' \in \mathbb{R}^\ell} \sum_{t=\ell+1}^T [\hat{\epsilon}_t(\rho, \hat{\boldsymbol{\pi}}_r) - c_1 \hat{\epsilon}_{t-1}(\rho, \hat{\boldsymbol{\pi}}_r) - \dots - c_\ell \hat{\epsilon}_{t-\ell}(\rho, \hat{\boldsymbol{\pi}}_r)]^2 \\ \widehat{\sigma}_\rho^2(\ell, \hat{\boldsymbol{\pi}}_r) &= \frac{1}{T-\ell} \sum_{t=\ell+1}^T [\hat{\epsilon}_t(\rho, \hat{\boldsymbol{\pi}}_r) - \hat{a}_{\rho,1}(\ell, \hat{\boldsymbol{\pi}}_r) \hat{\epsilon}_{t-1}(\rho, \hat{\boldsymbol{\pi}}_r) - \dots - \hat{a}_{\rho,\ell}(\ell, \hat{\boldsymbol{\pi}}_r) \hat{\epsilon}_{t-\ell}(\rho, \hat{\boldsymbol{\pi}}_r)]^2\end{aligned}$$

and $\widehat{\sigma}_\rho^2(0, \hat{\boldsymbol{\pi}}_r) = \frac{1}{T} \sum_{t=1}^T \hat{\epsilon}_t^2(\rho, \hat{\boldsymbol{\pi}}_r)$. Plug these estimates into $\mathbf{F}_\rho(q)$ and $\mathbf{S}_\rho(q)$ respectively,

$$\begin{aligned}\widehat{\mathbf{F}}_\rho(q, \hat{\boldsymbol{\pi}}_r) &= (\widehat{F}_{\rho,ij}(q, \hat{\boldsymbol{\pi}}_r), 1 \leq i, j \leq T) \\ \widehat{\mathbf{S}}_\rho(q, \hat{\boldsymbol{\pi}}_r) &= \text{diag}(\widehat{\sigma}_\rho^2(0, \hat{\boldsymbol{\pi}}_r), \dots, \widehat{\sigma}_\rho^2(q_T, \hat{\boldsymbol{\pi}}_r), \dots, \widehat{\sigma}_\rho^2(q_T, \hat{\boldsymbol{\pi}}_r)).\end{aligned}$$

The implied BIAM estimator is

$$\widehat{\Sigma}_\rho^{-1}(q, \hat{\boldsymbol{\pi}}_r) = \widehat{\mathbf{F}}_\rho'(q, \hat{\boldsymbol{\pi}}_r) \widehat{\mathbf{S}}_\rho^{-1}(q, \hat{\boldsymbol{\pi}}_r) \widehat{\mathbf{F}}_\rho(q, \hat{\boldsymbol{\pi}}_r). \quad (2.40)$$

(S3) Based on $\widehat{\Sigma}_\rho^{-1}(q, \hat{\boldsymbol{\pi}}_r)$, we construct

$$\widehat{\gamma}_{k,f}^\rho = [(\boldsymbol{\Delta}_\rho \mathbf{X}_k)' \widehat{\Sigma}_\rho^{-1}(q, \hat{\boldsymbol{\pi}}_r) (\boldsymbol{\Delta}_\rho \mathbf{X}_k)]^{-1} (\boldsymbol{\Delta}_\rho \mathbf{X}_k)' \widehat{\Sigma}_\rho^{-1}(q, \hat{\boldsymbol{\pi}}_r) (\boldsymbol{\Delta}_\rho \mathbf{y}),$$

and the FGLS-prewhitening break date estimator

$$\hat{k}_f^\rho = \arg \min_{k \in \Gamma_T} \text{SSR}\{k; \widehat{\gamma}_{k,f}^\rho, \boldsymbol{\Delta}_\rho' \widehat{\Sigma}_\rho^{-1}(q, \hat{\boldsymbol{\pi}}_r) \boldsymbol{\Delta}_\rho\}. \quad (2.41)$$

Plugging \hat{k}_f^ρ into $\widehat{\gamma}_{k,f}^\rho$, we also obtain the FGLS-prewhitening coefficient estimators.

The selection of QD parameter ρ remains flexible so far. For simplicity, we take $\rho = \hat{\rho}_r$, which gave satisfactory results for all our simulation studies. Finally, the same SSR in (2.41) shall be used in the construction of LR tests given in (2.31).

Remark 2.4

All the asymptotic results in previous sections still hold for the prewhitened estimators given a fixed $\rho \in (-1, 1)$. Since the proofs are similar and do not have much added value, we choose not to show them to avoid overburdened notation.

2.6 Simulations

In this section, we evaluate the finite sample performance of the proposed FGLS estimation and inference method. Two main themes are considered: (1) estimation efficiency; (2) coverage and length of confidence sets. We consider the DGP given in (2.1), where $\mu_0 = 0, \beta_0 = 0, (\mu_{b,0}, \beta_{b,0}) \in \{(0.5, 0.05), (1, 0.1)\}$, $k_0 = [\lambda_0 T]$, $\lambda_0 = 0.3^3$, $T \in \{150, 300\}$ and

$$\epsilon_t = \rho_0 \epsilon_{t-1} + \eta_t + \theta_0 \eta_{t-1}, \quad \eta_t \stackrel{i.i.d.}{\sim} \mathcal{N}(0, 0.1), \quad (2.42)$$

with $\rho_0 \in \{0.5, 0.7, 0.9, 0.95\}$ and $\theta_0 \in \{0.3, 0.6, 0.9\}$. Further details on implementations are as follow. First, we select the banding parameter q using the subsampling and risk-minimization method proposed by Bickel and Levina (2008). We use the same set of parameters as in Lin and Reuvers (2020b), except taking $[4.5 \times T^{0.15}]$ as an upper bound of possible banding parameters to adapt into our asymptotic framework. Second, for QD estimates, we take $r_m = \{0, 0.05, 0.1, \dots, 0.9, 0.95, 0.975\}$ with $m = 21$.⁴ Third, in the estimation of BIAM (Section 2.3.4), we obtain the first-stage residuals by plugging in the QD estimates. Finally, for Wald-type confidence intervals (Section 2.4.2) and LBI-based sets developed in Harvey and Leybourne (2015), we employ the Berk-type estimator of σ_ϵ^2 following the steps in Section 3.2 of Harvey and Leybourne (2015). More specifically, as suggested on their page 267, we construct σ_ϵ^2 using the residuals obtained by plugging QD estimates to improve the power of LBI tests. Similarly, we plug in FGLS estimates when

³Since different values of λ_0 , such as $\lambda_0 = 0.5, 0.7$, yield similar conclusions, we only report the outputs for $\lambda_0 = 0.3$.

⁴Harvey and Leybourne (2014) found that a finer grid with grid length 0.01 gives similar results.

the Wald-type intervals are constructed. The same procedure (MAIC) for lag length selection is adopted here as well. A presample of 200 observations is taken to eliminate the effects of initial values. The results are based on 10^4 Monte Carlo replications.

2.6.1 Discussions of the Simulation Results

First of all, we consider the efficiency of different LS break date estimators: (i) FGLS without prewhitening, (ii) improved FGLS with prewhitening (indicated by $\text{FGLS}_{\text{prewhiten}}$), (iii) OLS, (iv) QD. Considering the finite sample distributions of break date estimators in trending models are typically asymmetric and bimodal as reported in the literature, we adopt two indicators – mean squared error (MSE) and concentration probability (ConProb) – to describe both characteristics of centers and tails. ConProb measures the probability of break fraction estimates to lie in $[\lambda_0 - c, \lambda_0 + c]$ for a certain radius c . Our extensive, unreported simulations show the evidence that the ConProb for a small radius, say $c = 0.01$ or even smaller, is highly relevant to the performance of confidence sets. To show a better picture, we provide the outputs for both $c = 0.01$ and $c = 0.03$. The outputs are presented in Table 2.1 and are complemented by further results in the appendix.⁵ All the MSEs are expressed relative to the QD. A value above one shows better performance of the QD.

Some observations can be summarized as follows. First, notwithstanding the asymptotic equivalence in Proposition 2, notable improvements of FGLS exist in finite samples. Viewed holistically, the FGLS estimators improve the OLS and the QD for both MSE and ConProb, especially when the errors are more persistent. Second, the advantage of prewhitening is mostly revealed in ConProb under strong serial dependence, especially for larger breaks $(\mu_{b,0}, \beta_{b,0}) = (1, 0.1)$.

⁵As a benchmark, we also report the results of the infeasible GLS estimator in Section 2.2 in the appendix. The infeasible GLS estimator is constructed assuming the knowledge of true serial dependence.

Table 2.1: The empirical mean squared error (MSE) and concentration probability (ConProb) of break date estimators. The MSEs are expressed relative to QD. A value that is *larger* than 1 signifies that QD performs better. The concentration probability measures the probability of a break fraction estimator to lie in the interval $[\lambda_0 - c, \lambda_0 + c]$.

Efficiency			$(\mu_{b,0}, \beta_{b,0}) = (0.5, 0.05)$				$(\mu_{b,0}, \beta_{b,0}) = (1, 0.1)$			
	ρ_0	θ_0	FGLS _{prewhiten}	FGLS	OLS	QD	FGLS _{prewhiten}	FGLS	OLS	QD
$T = 150$										
MSE	0.5	0.6	0.851	0.679	0.713	1	0.643	0.552	1.233	1
	0.7	0.6	0.890	0.755	0.751	1	0.580	0.448	1.235	1
	0.95	0.6	0.909	0.871	0.691	1	0.474	0.661	1.100	1
	0.9	0.3	0.977	0.968	0.785	1	0.843	0.902	1.620	1
	0.9	0.6	0.905	0.845	0.710	1	0.524	0.585	1.155	1
	0.9	0.9	0.818	0.807	0.698	1	0.282	0.439	0.906	1
ConProb $c = 0.01$	0.5	0.6	0.279	0.290	0.192	0.227	0.830	0.834	0.580	0.738
	0.7	0.6	0.227	0.233	0.110	0.149	0.818	0.816	0.372	0.663
	0.95	0.6	0.193	0.159	0.046	0.096	0.765	0.643	0.115	0.485
	0.9	0.3	0.161	0.152	0.071	0.138	0.727	0.685	0.222	0.662
	0.9	0.6	0.196	0.180	0.058	0.106	0.782	0.717	0.153	0.537
	0.9	0.9	0.260	0.214	0.045	0.074	0.825	0.709	0.122	0.394
ConProb $c = 0.03$	0.5	0.6	0.397	0.421	0.340	0.336	0.880	0.890	0.727	0.809
	0.7	0.6	0.302	0.320	0.213	0.220	0.853	0.857	0.510	0.718
	0.95	0.6	0.232	0.200	0.107	0.136	0.780	0.663	0.192	0.511
	0.9	0.3	0.209	0.204	0.146	0.191	0.751	0.713	0.315	0.691
	0.9	0.6	0.242	0.229	0.126	0.153	0.800	0.739	0.245	0.571
	0.9	0.9	0.311	0.268	0.110	0.120	0.850	0.733	0.202	0.432
$T = 300$										
MSE	0.5	0.6	0.779	0.713	0.897	1	0.493	0.459	1.593	1
	0.7	0.6	0.781	0.571	0.854	1	0.398	0.355	1.865	1
	0.95	0.6	0.875	0.811	0.725	1	0.410	0.511	1.379	1
	0.9	0.3	1.044	0.802	0.830	1	0.885	0.714	2.026	1
	0.9	0.6	0.878	0.652	0.713	1	0.364	0.318	1.311	1
	0.9	0.9	0.661	0.542	0.694	1	0.072	0.121	0.974	1
ConProb $c = 0.01$	0.5	0.6	0.414	0.423	0.308	0.318	0.906	0.907	0.694	0.813
	0.7	0.6	0.351	0.365	0.166	0.226	0.893	0.895	0.470	0.728
	0.95	0.6	0.206	0.186	0.046	0.095	0.849	0.767	0.114	0.531
	0.9	0.3	0.193	0.204	0.082	0.176	0.784	0.781	0.258	0.730
	0.9	0.6	0.231	0.242	0.060	0.116	0.876	0.847	0.181	0.615
	0.9	0.9	0.401	0.362	0.048	0.079	0.961	0.900	0.134	0.452
ConProb $c = 0.03$	0.5	0.6	0.606	0.625	0.556	0.534	0.949	0.953	0.845	0.903
	0.7	0.6	0.501	0.527	0.372	0.407	0.929	0.932	0.681	0.826
	0.95	0.6	0.262	0.251	0.131	0.167	0.864	0.790	0.252	0.594
	0.9	0.3	0.294	0.327	0.201	0.288	0.837	0.838	0.461	0.798
	0.9	0.6	0.305	0.331	0.160	0.209	0.900	0.875	0.355	0.697
	0.9	0.9	0.459	0.436	0.135	0.157	0.971	0.916	0.290	0.551

Table 2.2: The coverage and length of nominal 0.95-level confidence sets with the break size $(\mu_{b,0}, \beta_{b,0})$.

Confidence set/interval			$(\mu_{b,0}, \beta_{b,0}) = (0.5, 0.05)$				$(\mu_{b,0}, \beta_{b,0}) = (1, 0.1)$			
	ρ_0	θ_0	C_{lbi}	$C_{lr,prewhiten}^*$	$C_{lr,prewhiten}$	$C_{wald,prewhiten}$	C_{lbi}	$C_{lr,prewhiten}^*$	$C_{lr,prewhiten}$	$C_{wald,prewhiten}$
$T = 150$										
Coverage	0.5	0.6	0.987	0.933	0.865	0.684	0.995	0.985	0.982	0.941
	0.7	0.6	0.979	0.929	0.819	0.691	0.988	0.983	0.980	0.938
	0.95	0.6	0.901	0.912	0.745	0.713	0.940	0.974	0.964	0.908
	0.9	0.3	0.958	0.899	0.692	0.718	0.983	0.968	0.955	0.911
	0.9	0.6	0.935	0.928	0.780	0.717	0.968	0.982	0.973	0.917
	0.9	0.9	0.928	0.942	0.854	0.713	0.954	0.985	0.977	0.933
Length	0.5	0.6	0.617	0.377	0.219	0.338	0.285	0.074	0.045	0.244
	0.7	0.6	0.807	0.569	0.308	0.437	0.422	0.126	0.066	0.325
	0.95	0.6	0.857	0.747	0.396	0.623	0.798	0.337	0.125	0.636
	0.9	0.3	0.889	0.680	0.309	0.565	0.725	0.262	0.099	0.481
	0.9	0.6	0.880	0.744	0.405	0.592	0.750	0.295	0.113	0.542
	0.9	0.9	0.887	0.778	0.510	0.601	0.794	0.305	0.128	0.596
$T = 300$										
Coverage	0.5	0.6	0.994	0.959	0.915	0.766	0.996	0.989	0.987	0.959
	0.7	0.6	0.989	0.954	0.896	0.798	0.994	0.989	0.988	0.968
	0.95	0.6	0.941	0.930	0.745	0.737	0.975	0.985	0.978	0.955
	0.9	0.3	0.977	0.912	0.729	0.750	0.992	0.982	0.973	0.958
	0.9	0.6	0.969	0.934	0.786	0.761	0.986	0.989	0.985	0.973
	0.9	0.9	0.962	0.960	0.909	0.777	0.982	0.993	0.992	0.988
Length	0.5	0.6	0.219	0.124	0.082	0.188	0.146	0.029	0.017	0.122
	0.7	0.6	0.296	0.198	0.117	0.255	0.179	0.037	0.021	0.163
	0.95	0.6	0.827	0.683	0.244	0.544	0.590	0.175	0.047	0.428
	0.9	0.3	0.634	0.452	0.164	0.398	0.306	0.098	0.042	0.267
	0.9	0.6	0.736	0.574	0.220	0.452	0.377	0.100	0.037	0.312
	0.9	0.9	0.790	0.633	0.338	0.480	0.452	0.077	0.027	0.347

Now we compare the coverage and length of the confidence sets/intervals.⁶ As benchmarks, we compare the suggested LR-based methods to the Wald-type intervals (Section 2.4.2) and the LBI-based confidence sets by Harvey and Leybourne (2015). For the sets/intervals developed in this paper, we only report the results (indicated by the subscript ‘prewhiten’) based on the prewhitening FGLS to save space. The interested reader is referred to the appendix for the full results including those without prewhitening. The outputs are reported in Tables 2.2.

Overall, interval $C_{lr,prewhiten}^*$, which is the combination of $FGLS_{prewhiten}$ and the LR-based interval C_{lr}^* , works the most satisfactorily on the grounds of

⁶Traditionally, we compute the length (i) using the cardinality of a set, (ii) then normalizing it by the cardinality of Γ_T , (iii) and averaging over replications.

both coverage and length. The Wald-type intervals and the LR-based sets have the shortest lengths, but showing severe under-coverage for small breaks. On the other hand, C_{lbi} has good coverage but is the most lengthy. The interval $C_{lr,prewhiten}^*$ has well-pleasing coverage in most scenarios, yet reducing the length of C_{lbi} by roughly 20% up to 80% (e.g. $(\mu_{b,0}, \beta_{b,0}) = (1, 0.1)$, $\rho_0 = \theta_0 = 0.9$).⁷ In the case of $(\mu_{b,0}, \beta_{b,0}) = (1, 0.1)$, the confidence set $C_{lr,prewhiten}$ has the optimal performance for its ideal coverage and the shortest length, followed closely by $C_{lr,prewhiten}^*$. Furthermore, by fixing the value of the MA coefficient θ_0 and increasing the AR coefficient ρ_0 , the coverages of $C_{lr,prewhiten}^*$ and $C_{lr,prewhiten}$ show certain uniform property across $I(0)/I(1)$ dichotomy, especially for $(\mu_{b,0}, \beta_{b,0}) = (1, 0.1)$. Although our theory assumes weakly stationary errors, these methods can be potentially applied in the models with $I(1)$ errors. Since $C_{lr,prewhiten}^*$ can have slight under-coverage for small T and small break magnitudes, we recommend researchers to consider both $C_{lr,prewhiten}^*$ and C_{lbi} . If the length of $C_{lr,prewhiten}^*$ is much shorter than C_{lbi} , for instance, it is about 1/3 for $(\mu_{b,0}, \beta_{b,0}) = (1, 0.1)$, then $C_{lr,prewhiten}^*$ can be a reasonable choice.

In Figure 2.2, we further investigate how the performance of confidence sets/intervals is affected by the magnitudes of break size. In the upper panel, we fix the slope change $\beta_{b,0} = 0.05$ and increase the level break $\mu_{b,0}$ from 0 to 1.2. Similarly, in the lower panel, the level change $\mu_{b,0} = 0.5$ is fixed while $\beta_{b,0}$ is taken from 0 to 0.12. The remaining parameters remained unchanged in both panels, namely $\lambda_0 = 0.3$, $T = 150$, $\rho_0 = 0.7$ and $\theta_0 = 0.6$. The coverage of all the methods improves with increasing break sizes, except C_{lbi} which is nearly invariant. All of them have shorter length when $\beta_{b,0}$ increases. Interestingly, the length of C_{lbi} first increases and then remains at a high level even with a large $\mu_{b,0}$. This is probably due to the nonmonotonic power problem of a local asymptotic framework as discussed.

⁷The advantage of length of $C_{lr,prewhiten}^*$ compared to C_{lbi} can be more significant if the median length is used instead of the average length. In our unreported simulations, the distribution of length of $C_{lr,prewhiten}^*$ is likely to be right-skewed, while C_{lbi} tends to be left-skewed.

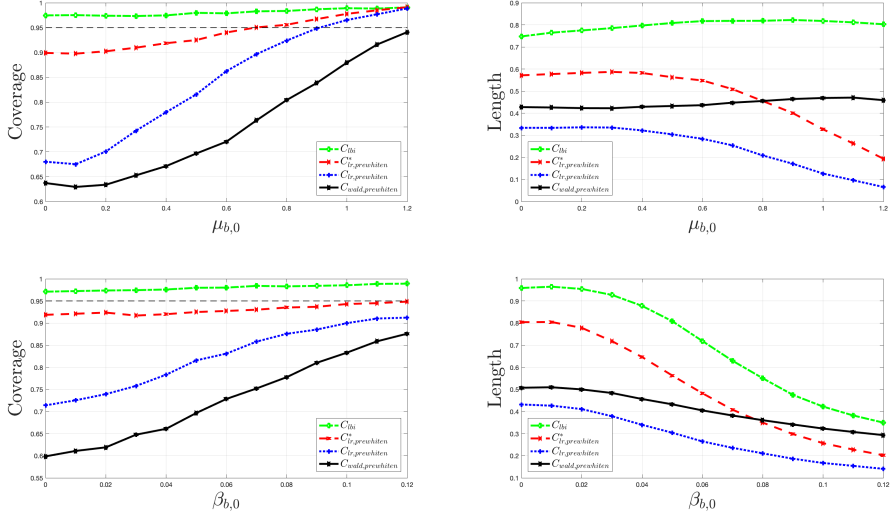


Figure 2.2: The coverage (left column) and length (right column) of confidence sets/intervals with varying break sizes. In the upper panel, we take $(\mu_{b,0}, \beta_{b,0}) = (\mu_{b,0}, 0.05)$, where $\mu_{b,0} = 0, 0.1, \dots, 1.2$. In the lower panel, we take $(\mu_{b,0}, \beta_{b,0}) = (0.5, \beta_{b,0})$, where $\beta_{b,0} = 0, 0.01, \dots, 0.12$.

2.7 Conclusion

We proposed a (feasible) GLS estimation and inference framework in models with breaks in level and/or trend slope. The suggested FGLS methods reduce the mean squared error and improve the concentration probability of existing approaches in finite samples. The confidence sets/intervals based on inverting multiple likelihood ratio tests inherit the estimation efficiency of FGLS, and show satisfactory coverage and length properties in various situations. There are several possible extensions for future research. For instance, it is interesting to generalize the methods to allow an unknown number of (multiple) breaks and nonstationary errors.

Appendix 2.A Preliminary Results for Infeasible GLS

Lemma 2.1

Let $A_q(L) = 1 - \sum_{j=1}^q a_j(q)L^j$ be the filter obtained by MCD with q lags. By the Beveridge-Nelson (BN) decomposition, we can express $A_q(L)$ as $A_q(L) = A_q(1) + (1-L)\tilde{A}_q(L)$, where $\tilde{A}_q(L) = \sum_{j=1}^q \tilde{a}_j(q)L^{j-1}$ and $\tilde{a}_j(q) = \sum_{i=j}^q a_i(q)$. Let $B_q(L) = A_q(L)A(L)^{-1} = \sum_{j=0}^{\infty} b_j(q)L^j$, where $A(L) = 1 - \sum_{j=1}^{\infty} a_j L^j$ is given in (2.3). Suppose $q = q_T \rightarrow \infty$ as $T \rightarrow \infty$, under Assumption 2.3,

$$(i) \lim_{T \rightarrow \infty} A_q(1) = A(1) \text{ and } \lim_{T \rightarrow \infty} \sum_{j=1}^q j |\tilde{a}_j(q)| < \infty;$$

$$(ii) \lim_{T \rightarrow \infty} \sum_{j=1}^{\infty} j |b_j(q)| < \infty.$$

Proof Assumption 2.3 enables us to apply the generalized Baxter's inequality, see Theorem 6.6.12 of Hannan and Deistler (2012). First, $|A_q(1) - A(1)| \leq \sum_{j=1}^q |a_j(q) - a_j| + \sum_{j=q+1}^{\infty} |a_j| \leq C \sum_{j=q+1}^{\infty} |a_j| \rightarrow 0$ as $T \rightarrow \infty$. Second, by interchanging the order of summation, we get

$$\sum_{j=1}^q j |\tilde{a}_j(q)| \leq \sum_{i=1}^q \sum_{j=1}^i j |a_i(q)| = \frac{1}{2} \sum_{j=1}^q (j^2 + j) |a_j(q)|. \quad (2.43)$$

By the generalized Baxter's inequality again, we have $\sum_{j=1}^q j^2 |a_j(q)| \leq \sum_{j=1}^q j^2 |a_j(q) - a_j| + \sum_{j=1}^q j^2 |a_j| \leq C \sum_{j=1}^{\infty} j^2 |a_j| < \infty$. Then the second result can be seen by letting $T \rightarrow \infty$ in the both sides of (2.43).

Next we consider the summability of the coefficients of $B_q(L)$. For convenience, let $C(L) = A(L)^{-1} = \sum_{j=0}^{\infty} c_j L^j$ with $c_0 = 1$. Assumption 2.3 implies $\sum_{j=0}^{\infty} j^2 |c_j| < \infty$. Note that

$$\begin{aligned} B_q(L) &= \left[1 - \sum_{j=1}^q a_j(q)L^j \right] \left[\sum_{j=0}^{\infty} c_j L^j \right] = \sum_{j=0}^{\infty} c_j L^j - \sum_{j=0}^{\infty} \sum_{i=1}^q c_j a_i(q) L^{i+j} \\ &= \sum_{j=0}^{\infty} c_j L^j - \sum_{j=1}^{\infty} \sum_{i=1}^{\min\{q,j\}} c_{j-i} a_i(q) L^j = 1 + \sum_{j=1}^{\infty} \left[c_j - \sum_{i=1}^{\min\{q,j\}} c_{j-i} a_i(q) \right] L^j. \end{aligned}$$

By construction, $b_0(q) = 1$, $b_j(q) = c_j - \sum_{i=1}^{\min\{q,j\}} c_{j-i}a_i(q)$ for $j \geq 1$. As $T \rightarrow \infty$,

$$\begin{aligned}
\sum_{j=1}^{\infty} j|b_j(q)| &\leq \sum_{j=1}^{\infty} j|c_j| + \sum_{j=1}^q \sum_{i=1}^j j|c_{j-i}a_i(q)| + \sum_{j=q+1}^{\infty} \sum_{i=1}^q j|c_{j-i}a_i(q)| \\
&= \sum_{j=1}^{\infty} j|c_j| + \sum_{i=1}^q |a_i(q)| \sum_{j=i}^{\infty} j|c_{j-i}| \\
&= \sum_{j=1}^{\infty} j|c_j| + \sum_{j=1}^{\infty} j|c_j| \sum_{j=1}^q |a_j(q)| + \sum_{j=0}^{\infty} |c_j| \sum_{j=1}^q j|a_j(q)| < \infty,
\end{aligned} \tag{2.44}$$

where (2.44) is obtained by interchanging the order of summation. \blacksquare

With a slight abuse of notation, the premultiplication by a lag polynomial in the subsequent proofs should be understood as an operation on an element of a time series. For instance, $A_q(L) \cdot 1$ means we are looking at $A_q(L) \cdot c_t$, where $c_t = 1$ for $t = 1, 2, \dots, T$. Similarly, $A_q(L)\mathbf{1}_{\{t>k\}}$ stands for $A_q(L)c_t$ with $c_t = \mathbf{1}_{\{t>k\}}$.

Lemma 2.2

For $t = 1, 2, \dots, T$, $k, k_1, k_2 \in \Gamma_T$, suppose Assumption 2.3 holds, then $\sup_{t \in \mathbb{R}} \mathbb{E}|\tilde{A}_q(L)\epsilon_t|^4 < \infty$ and

- (i) $A_q(L) \cdot 1 = A_q(1)$;
- (ii) $\left[A_q(L)\frac{t}{T}\right]^j = \left[A_q(1)\frac{t}{T}\right]^j + O(T^{-1})$, where $j \in \mathbb{N}^+$;
- (iii) $A_q(L)\mathbf{1}_{\{t>k\}} = A_q(1)\mathbf{1}_{\{t>k\}} + \tilde{A}_q(L)\mathbf{1}_{\{t=k+1\}}$, where the second term in the RHS is bounded by $|\tilde{A}_q(L)\mathbf{1}_{\{t=k+1\}}| \leq \sum_{j=1}^q |\tilde{a}_j(q)|\mathbf{1}_{\{t=k+j\}}$;
- (iv) $A_q(L)(t-k)\mathbf{1}_{\{t>k\}} = A_q(1)(t-k)\mathbf{1}_{\{t>k\}} + \tilde{A}_q(L)\mathbf{1}_{\{t>k\}}$. Moreover, for $k_1 \neq k_2$,
 $A_q(L)(t-k_1)\mathbf{1}_{\{t>k_2\}} = A_q(1)(t-k_1)\mathbf{1}_{\{t>k_2\}} + \tilde{A}_q(L)\mathbf{1}_{\{t>k_2\}} + (k_2-k_1)\tilde{A}_q(L)\mathbf{1}_{\{t=k_2+1\}}$,
where $|\tilde{A}_q(L)\mathbf{1}_{\{t>k_2\}}| \leq C\mathbf{1}_{\{t>k_2\}}$.
- (v) $[A_q(L) \cdot 1][A_q(L)\epsilon_t] = A_q(1)^2\epsilon_t + A_q(1)\Delta[\tilde{A}_q(L)\epsilon_t]$;

$$(vi) \left[A_q(L) \frac{t}{T} \right] [A_q(L) \epsilon_t] = A_q(1)^2 \frac{t}{T} \epsilon_t + A_q(1) \Delta \left[\frac{t}{T} (\tilde{A}_q(L) \epsilon_t) \right] + O_p(T^{-1}).$$

Proof First of all, by Minkowski's inequality, see e.g. Stein (1970, A.1), and the law of iterated expectation, Assumption 2.3 implies $\sup_{t \in \mathbb{R}} \mathbb{E} |\epsilon_t|^4 < \infty$. Again by Minkowski's inequality, using Lemma 2.1 (i),

$$\sup_{t \in \mathbb{R}} \mathbb{E} |\tilde{A}_q(L) \epsilon_t|^4 = \sup_{t \in \mathbb{R}} \mathbb{E} \left| \sum_{j=1}^q \tilde{a}_j(q) \epsilon_{t-j+1} \right|^4 \leq \left[\sum_{j=1}^q |\tilde{a}_j(q)| \right]^4 \sup_{t \in \mathbb{R}} \mathbb{E} |\epsilon_t|^4 < \infty. \quad (2.45)$$

Now we consider the six items in the lemma. By the BN decomposition, we immediately obtain Items (i) and (v). Part (ii) is obtained by further using the binomial expansion. For Item (iii), we can write

$$A_q(L) \mathbf{1}_{\{t > k\}} = A_q(1) \mathbf{1}_{\{t > k\}} + \tilde{A}_q(L) (1 - L) \mathbf{1}_{\{t > k\}} = A_q(1) \mathbf{1}_{\{t > k\}} + \tilde{A}_q(L) \mathbf{1}_{\{t = k+1\}}$$

which is the first part of (iii). Moreover, the second term in the right hand side above is bounded by $\left| \sum_{j=1}^q \tilde{a}_j(q) \mathbf{1}_{\{t=k+j\}} \right| \leq \sum_{j=1}^q |\tilde{a}_j(q)| \mathbf{1}_{\{t=k+j\}}$. For Item (iv), we only show the case $k_1 \neq k_2$. The BN decomposition implies

$$\begin{aligned} A_q(L)(t - k_1) \mathbf{1}_{\{t > k_2\}} &= A_q(1)(t - k_1) \mathbf{1}_{\{t > k_2\}} + \tilde{A}_q(L)(1 - L)(t - k_1) \mathbf{1}_{\{t > k_2\}} \\ &= A_q(1)(t - k_1) \mathbf{1}_{\{t > k_2\}} + \tilde{A}_q(L) [(t - k_1) \mathbf{1}_{\{t > k_2\}} - (t - 1 - k_1) \mathbf{1}_{\{t > k_2+1\}}] \\ &= A_q(1)(t - k_1) \mathbf{1}_{\{t > k_2\}} + \tilde{A}_q(L)(t - k_1) \mathbf{1}_{\{t = k_2+1\}} + \tilde{A}_q(L) \mathbf{1}_{\{t > k_2+1\}} \\ &= A_q(1)(t - k_1) \mathbf{1}_{\{t > k_2\}} + \tilde{A}_q(L)(k_2 - k_1 + 1) \mathbf{1}_{\{t = k_2+1\}} + \tilde{A}_q(L) \mathbf{1}_{\{t > k_2+1\}} \\ &= A_q(1)(t - k_1) \mathbf{1}_{\{t > k_2\}} + \tilde{A}_q(L) \mathbf{1}_{\{t > k_2\}} + (k_2 - k_1) \tilde{A}_q(L) \mathbf{1}_{\{t = k_2+1\}} \end{aligned}$$

where the third equality uses $\mathbf{1}_{\{t > k_2\}} = \mathbf{1}_{\{t = k_2+1\}} + \mathbf{1}_{\{t > k_2+1\}}$. Now by the fact that $\mathbf{1}_{\{t > k_2\}} \geq \mathbf{1}_{\{t > k_2+1\}}$,

$$|\tilde{A}_q(L) \mathbf{1}_{\{t > k_2\}}| = \left| \sum_{j=1}^q \tilde{a}_j(q) \mathbf{1}_{\{t > k_2+j-1\}} \right| \leq \sum_{j=1}^q |\tilde{a}_j(q)| \mathbf{1}_{\{t > k_2+j-1\}} \leq C \mathbf{1}_{\{t > k_2\}}, \quad (2.46)$$

and Part (iv) is then obtained. Finally, we note that

$$\frac{t}{T} \Delta [\tilde{A}_q(L) \epsilon_t] = \Delta \left[\frac{t}{T} (\tilde{A}_q(L) \epsilon_t) \right] - \frac{1}{T} [\tilde{A}_q(L) \epsilon_{t-1}] = \Delta \left[\frac{t}{T} (\tilde{A}_q(L) \epsilon_t) \right] + O_p(T^{-1}). \quad (2.47)$$

Then (vi) is straightforward by the BN decomposition and (2.47). \blacksquare

Lemma 2.3

Let $W(\cdot)$ be a standard Brownian motion defined on $[0, 1]$. Under Assumptions 2.1 to 2.4, $\forall \tau \in [0, 1]$,

- (i) $T^{-1/2} \sigma^{-2}(q) \sum_{t=1}^{[\tau T]} [A_q(L) \cdot 1][A_q(L)\epsilon_t] \xrightarrow{\mathcal{D}} \sigma_\epsilon^{-1} W(\tau);$
- (ii) $T^{-3/2} \sigma^{-2}(q) \sum_{t=1}^{[\tau T]} [A_q(L) \cdot t][A_q(L)\epsilon_t] \xrightarrow{\mathcal{D}} \sigma_\epsilon^{-1} \int_0^\tau r dW(r);$
- (iii) $v_T \sigma^{-2}(q) \sum_{t=k_0+1}^{k_0+[\tau v_T^{-2}]} [A_q(L) \cdot 1][A_q(L)\epsilon_t] \xrightarrow{\mathcal{D}} \sigma_\epsilon^{-1} W(\tau);$
- (iv) $v_T^3 \sigma^{-2}(q) \sum_{t=k_0+1}^{k_0+[\tau v_T^{-2}]} [A_q(L)(t - k_0)][A_q(L)\epsilon_t] \xrightarrow{\mathcal{D}} \sigma_\epsilon^{-1} \int_0^\tau r dW(r).$

Proof (i) Note that $\sigma^2(q)A_q(1)^{-2} \rightarrow \sigma^2 A(1)^{-2} = \sigma_\epsilon^2$. By Lemma 2.2 (v), we have

$$\begin{aligned} T^{-1/2} \sigma^{-2}(q) \sum_{t=1}^{[\tau T]} [A_q(L) \cdot 1][A_q(L)\epsilon_t] \\ = \sigma^{-2}(q) A_q(1)^2 T^{-1/2} \sum_{t=1}^{[\tau T]} \epsilon_t + \sigma^{-2}(q) A_q(1) T^{-1/2} \tilde{A}_q(L) (\epsilon_{[\tau T]} - \epsilon_0). \end{aligned}$$

By a functional central limit theorem (FCLT) for linear processes (Theorem 3.15 of Phillips and Solo (1992)), $T^{-1/2} \sum_{t=1}^{[\tau T]} \epsilon_t \xrightarrow{\mathcal{D}} \sigma_\epsilon W(\tau)$. Since the second term is $O_p(T^{-1/2})$, we obtain the result. (ii) For the second result, Lemma 2.2 (vi) gives

$$T^{-3/2} \sigma^{-2}(q) \sum_{t=1}^{[\tau T]} [A_q(L) \cdot t][A_q(L)\epsilon_t] = \sigma^{-2}(q) A_q(1)^2 T^{-1/2} \sum_{t=1}^{[\tau T]} \frac{t}{T} \epsilon_t + O_p(T^{-1/2}).$$

Then Part (ii) follows from standard textbook arguments, see, e.g., Section 17.5 of Hamilton (1994). (iii) Similarly as Theorem 1 of Bai (1994),

$$v_T \sigma^{-2}(q) \sum_{t=k_0+1}^{k_0+[\tau v_T^{-2}]} [A_q(L) \cdot 1][A_q(L)\epsilon_t] = \sigma^{-2}(q) A_q(1)^2 v_T \sum_{t=1}^{[\tau v_T^{-2}]} \epsilon_{k_0+t} + O_p(v_T).$$

By the FCLT, we obtain (iii). Part (iv) is similar to (ii) and is thus omitted. ■

We introduce a crucial decomposition of $\text{SSR}(k_0; \hat{\gamma}_{k_0}, \Sigma^{-1}(q)) - \text{SSR}(k; \hat{\gamma}_k, \Sigma^{-1}(q))$ that reduces computational effort in the proofs later.

Lemma 2.4

The difference $SSR(k_0; \hat{\gamma}_{k_0}, \Sigma^{-1}(q)) - SSR(k; \hat{\gamma}_k, \Sigma^{-1}(q))$ can be decomposed into three key terms:

$$SSR(k_0; \hat{\gamma}_{k_0}, \Sigma^{-1}(q)) - SSR(k; \hat{\gamma}_k, \Sigma^{-1}(q)) = -(XX)_k - 2(XE)_k - (EE)_k, \quad (2.48)$$

where

$$\begin{aligned} (XX)_k &= \mathbf{d}'_k \left\{ \Sigma^{-1}(q) - \Sigma^{-1}(q) \mathbf{X}_k \left[\mathbf{X}'_k \Sigma^{-1}(q) \mathbf{X}_k \right]^{-1} \mathbf{X}'_k \Sigma^{-1}(q) \right\} \mathbf{d}_k, \\ (XE)_k &= \mathbf{d}'_k \Sigma^{-1}(q) \varepsilon - \mathbf{d}'_k \Sigma^{-1}(q) \mathbf{X}_k \left[\mathbf{X}'_k \Sigma^{-1}(q) \mathbf{X}_k \right]^{-1} \mathbf{X}'_k \Sigma^{-1}(q) \varepsilon, \\ (EE)_k &= \varepsilon' \Sigma^{-1}(q) \mathbf{X}_{k_0} \left[\mathbf{X}'_{k_0} \Sigma^{-1}(q) \mathbf{X}_{k_0} \right]^{-1} \mathbf{X}'_{k_0} \Sigma^{-1}(q) \varepsilon \\ &\quad - \varepsilon' \Sigma^{-1}(q) \mathbf{X}_k \left[\mathbf{X}'_k \Sigma^{-1}(q) \mathbf{X}_k \right]^{-1} \mathbf{X}'_k \Sigma^{-1}(q) \varepsilon, \end{aligned}$$

with

$$\mathbf{d}_k = (d_{k,1}, \dots, d_{k,T})' = [\mathbf{X}_{k_0} - \mathbf{X}_k - (k - k_0)\mathbf{Z}_k] \gamma_0 \quad (2.49)$$

and \mathbf{Z}_k is a $T \times 4$ matrix that the first three columns are $\mathbf{0}$'s and the last column takes the third column of \mathbf{X}_k , that is, $\mathbf{Z}_k = (z_{k,1}, \dots, z_{k,T})'$, $z_{k,t} = (0, 0, 0, \mathbf{1}_{\{t > k\}})'$. The entries of \mathbf{d}_k can be explicitly given as:

$$d_{k,t} = \text{sgn}(k - k_0) [\mu_{b,0} + \beta_{b,0}(t - k_0)] \mathbf{1}_{\{\min\{k, k_0\} + 1 \leq t \leq \max\{k, k_0\}\}}. \quad (2.50)$$

Proof Using the orthogonality property

$$\left\{ \Sigma^{-1}(q) - \Sigma^{-1}(q) \mathbf{X}_k \left[\mathbf{X}'_k \Sigma^{-1}(q) \mathbf{X}_k \right]^{-1} \mathbf{X}'_k \Sigma^{-1}(q) \right\} \mathbf{Z}_k = \mathbf{0},$$

the results follow by straightforward linear algebra, hence are omitted here. ■

For any $M_0 > 0$, we first define two subsets of potential break points

$$\begin{aligned} V_T(M_0) &= \{k : |k - k_0| > M_0 v_T^{-2}, k \in \Gamma_T\} \\ K_T(M_0) &= \{k : |k - k_0| \leq M_0 v_T^{-2}, k \in \Gamma_T\}. \end{aligned} \quad (2.51)$$

We now derive the asymptotic orders of some crucial quantities that are related to the three key objects in (2.48).

Lemma 2.5

Denote $\lambda = k/T$, and recall that $\mathbf{D}_T = \text{diag}(T, T^3, T, T^3)$. Under Assumptions 2.1, 2.2 and 2.4,

$$\left[\mathbf{D}_T^{-1/2} \mathbf{X}'_k \boldsymbol{\Sigma}^{-1}(q) \mathbf{X}_k \mathbf{D}_T^{-1/2} \right]^{-1} = \sigma^2(q) A_q(1)^{-2} \boldsymbol{\Omega}_\lambda^{-1} + O(T^{-1}), \quad (2.52)$$

for any $k \in \Gamma_T$, where $\boldsymbol{\Omega}_\lambda^{-1}$ is defined in (2.20) in Theorem 2.2, and $O(T^{-1})$ is uniform on Γ_T . Furthermore,

$$\sup_{k \in K_T(M_0)} \left\| \left[\mathbf{D}_T^{-1/2} \mathbf{X}'_k \boldsymbol{\Sigma}^{-1}(q) \mathbf{X}_k \mathbf{D}_T^{-1/2} \right]^{-1} - \sigma_\epsilon^2 \boldsymbol{\Omega}_{\lambda_0}^{-1} \right\| = o(1). \quad (2.53)$$

Proof Note that $\mathbf{D}_T^{-1/2} \mathbf{X}'_k \boldsymbol{\Sigma}^{-1}(q) \mathbf{X}_k \mathbf{D}_T^{-1/2} = \mathbf{D}_T^{-1/2} [\mathbf{F}(q) \mathbf{X}_k]' \mathbf{S}^{-1}(q) [\mathbf{F}(q) \mathbf{X}_k] \mathbf{D}_T^{-1/2}$. The premultiplication of \mathbf{X}_k by $\mathbf{F}(q)$ applies filters to the entries of \mathbf{X}_k , and $\mathbf{S}^{-1}(q)$ implies weightings. The filter and scaling factor are precisely $A_q(L)$ and $\sigma^{-2}(q)$, respectively, for the components in the last $T - q$ rows of \mathbf{X}_k . Under Assumption 2.4, it is not hard to show that the first $q \ll T$ rows do not play asymptotic roles. Since the rigorous proof only requires a slight modification with more cumbersome notation, we may assume (without loss of generality) that $A_q(L)$ and $\sigma^{-2}(q)$ applies for all $t = 1, 2, \dots, T$. Employing the BN decomposition for $A_q(L) \mathbf{x}_{k,t}$ element-wise, and using the results (i) - (iv) in Lemma 2.2, we obtain

$$\begin{aligned} \mathbf{D}_T^{-1/2} \mathbf{X}'_k \boldsymbol{\Sigma}^{-1}(q) \mathbf{X}_k \mathbf{D}_T^{-1/2} &= \sigma^{-2}(q) \sum_{t=1}^T \mathbf{D}_T^{-1/2} [A_q(L) \mathbf{x}_{k,t}] [A_q(L) \mathbf{x}_{k,t}]' \mathbf{D}_T^{-1/2} \\ &= \sigma^{-2}(q) A_q(1)^2 \frac{1}{T} \sum_{t=1}^T \begin{pmatrix} 1 & \frac{t}{T} & \mathbf{1}_{\{t>k\}} & \frac{t-k}{T} \mathbf{1}_{\{t>k\}} \\ \frac{t}{T} & \frac{t^2}{T^2} & \frac{t}{T} \mathbf{1}_{\{t>k\}} & \frac{t(t-k)}{T^2} \mathbf{1}_{\{t>k\}} \\ \mathbf{1}_{\{t>k\}} & \frac{t}{T} \mathbf{1}_{\{t>k\}} & \mathbf{1}_{\{t>k\}} & \frac{t-k}{T} \mathbf{1}_{\{t>k\}} \\ \frac{t-k}{T} \mathbf{1}_{\{t>k\}} & \frac{t(t-k)}{T^2} \mathbf{1}_{\{t>k\}} & \frac{t-k}{T} \mathbf{1}_{\{t>k\}} & \frac{(t-k)^2}{T^2} \mathbf{1}_{\{t>k\}} \end{pmatrix} + O(T^{-1}) \end{aligned} \quad (2.54)$$

$$= \sigma^{-2}(q) A_q(1)^2 \boldsymbol{\Omega}_\lambda + O(T^{-1}) \quad (2.55)$$

where

$$\mathbf{\Omega}_\lambda = \begin{pmatrix} 1 & \frac{1}{2} & 1 - \lambda & \frac{(1-\lambda)^2}{2} \\ \frac{1}{2} & \frac{1}{3} & \frac{1-\lambda^2}{2} & \frac{(1-\lambda)^2(2+\lambda)}{6} \\ 1 - \lambda & \frac{1-\lambda^2}{2} & 1 - \lambda & \frac{(1-\lambda)^2}{2} \\ \frac{(1-\lambda)^2}{2} & \frac{(1-\lambda)^2(2+\lambda)}{6} & \frac{(1-\lambda)^2}{2} & \frac{(1-\lambda)^3}{3} \end{pmatrix}. \quad (2.56)$$

The $O(T^{-1})$ term in (2.54) originates from the BN decomposition, containing the parts related to lag polynomial $\tilde{A}_q(L)$. The $O(T^{-1})$ term in (2.55) includes additional errors between the scaled partial sum in (2.54) and $\mathbf{\Omega}_\lambda$. Take the $(2, 4)_{th}$ element $\frac{1}{T} \sum_{t=1}^T \frac{t(t-k)}{T^2} \mathbf{1}_{\{t>k\}}$ as an example. By calculation, we have

$$\begin{aligned} \frac{1}{T} \sum_{t=1}^T \frac{t(t-k)}{T^2} \mathbf{1}_{\{t>k\}} &= \frac{1}{T} \sum_{t=k+1}^T \frac{t(t-k)}{T^2} \\ &= \frac{1}{6} \frac{(T-k)^2(2T+k)}{T^3} + \frac{1}{6} \frac{(T-k)(3T+1)}{T^3} = \frac{(1-\lambda)^2(2+\lambda)}{6} + O(T^{-1}). \end{aligned}$$

It is not hard to see that both the $O(T^{-1})$ terms in (2.54) and (2.55) are uniform on Γ_T .

For any $\lambda \in [\delta_1, 1 - \delta_2]$, it is easy to verify that all the leading principal minors of $\mathbf{\Omega}_\lambda$ are positive, hence $\mathbf{\Omega}_\lambda$ is positive definite and then invertible. By taking the inverse, we have (2.52). Finally, we obtain (2.53) by noting $\sigma^2(q)A_q(1)^{-2} \rightarrow \sigma^2 A(1)^{-2} = \sigma_\epsilon^2$ and $\lambda \rightarrow \lambda_0$ as $T \rightarrow \infty$. ■

Lemma 2.6

Denote $\delta_k = |k - k_0|$. By the definition of $d_{k,t}$ in (2.50), for any $k \in \Gamma_T$, we can express the following (*partial, squared and weighted*) summation quantities related to $d_{k,t}$ as

$$\begin{aligned} \sum_{t=1}^T d_{k,t} &= \sum_{t=\min\{k, k_0\}+1}^{\max\{k, k_0\}} d_{k,t} = \text{sgn}(k - k_0) d_{\mu,0} \delta_k v_T + \frac{1}{2} d_{\beta,0} \delta_k^2 v_T^3 + R_p(\delta_k, v_T) \\ \sum_{t=1}^T d_{k,t}^2 &= \sum_{t=\min\{k, k_0\}+1}^{\max\{k, k_0\}} d_{k,t}^2 = \frac{1}{3} d_{\beta,0}^2 \delta_k^3 v_T^6 + \text{sgn}(k - k_0) d_{\mu,0} d_{\beta,0} \delta_k^2 v_T^4 + d_{\mu,0}^2 \delta_k v_T^2 + R_s(\delta_k, v_T) \\ \sum_{t=1}^T t d_{k,t} &= \sum_{t=\min\{k, k_0\}+1}^{\max\{k, k_0\}} t d_{k,t} = \text{sgn}(k - k_0) \frac{1}{2} d_{\mu,0} \delta_k (k + k_0) v_T + \frac{1}{6} d_{\beta,0} \delta_k^2 (2k + k_0) v_T^3 + R_w(\delta_k, v_T) \end{aligned} \quad (2.57)$$

where R_p , R_s and R_w are some functions of k , k_0 and v_T given in the proof.

Suppose Assumptions 2.1 and 2.2 hold. If $d_{\beta,0} \neq 0$, for any $M_0 > 0$ and T ,

$$\begin{aligned}
\sup_{\{\delta_k: \delta_k > M_0 v_T^{-2}\}} |(\delta_k v_T^3)^{-1} R_p(\delta_k, v_T)| &\leq C, & \sup_{\{\delta_k: \delta_k \leq M_0 v_T^{-2}\}} |R_p(\delta_k, v_T)| &\leq C v_T, \\
\sup_{\{\delta_k: \delta_k > M_0 v_T^{-2}\}} |(\delta_k^2 v_T^6)^{-1} R_s(\delta_k, v_T)| &\leq C, & \sup_{\{\delta_k: \delta_k \leq M_0 v_T^{-2}\}} |R_s(\delta_k, v_T)| &\leq C v_T^2, \\
\sup_{\{\delta_k: \delta_k > M_0 v_T^{-2}\}} |(T \delta_k v_T^3)^{-1} R_w(\delta_k, v_T)| &\leq C, & \sup_{\{\delta_k: \delta_k \leq M_0 v_T^{-2}\}} |R_w(\delta_k, v_T)| &\leq C T v_T^{-1}.
\end{aligned} \tag{2.58}$$

If $d_{\beta,0} = 0$, then $R_p = R_s \equiv 0$, and

$$\sup_{\{\delta_k: \delta_k > M_0 v_T^{-2}\}} |(\delta_k v_T)^{-1} R_w(\delta_k, v_T)| \leq C, \quad \sup_{\{\delta_k: \delta_k \leq M_0 v_T^{-2}\}} |R_w(\delta_k, v_T)| \leq C v_T^{-1}.$$

Proof Direct calculation yields (2.57) with

$$\begin{aligned}
R_p(\delta_k, v_T) &= \text{sgn}(k - k_0) \frac{1}{2} d_{\beta,0} \delta_k v_T^3 \\
R_s(\delta_k, v_T) &= \text{sgn}(k - k_0) \frac{1}{2} d_{\beta,0}^2 \delta_k^2 v_T^6 + \frac{1}{6} d_{\beta,0}^2 \delta_k v_T^6 + d_{\mu,0} d_{\beta,0} \delta_k v_T^4 \\
R_w(\delta_k, v_T) &= \text{sgn}(k - k_0) \frac{1}{2} d_{\mu,0} \delta_k v_T + \frac{1}{6} d_{\beta,0} (3\delta_k^2 + \text{sgn}(k - k_0) \delta_k (3k_0 + 1)) v_T^3.
\end{aligned}$$

The results for R_p and R_s are straightforward. The results for R_w follow from the assumption $k_0 = [\lambda_0 T]$. \blacksquare

Lemma 2.7

Under Assumptions 2.1, 2.2 and 2.4, for any $k \in \Gamma_T$,

$$\begin{aligned}
&D_T^{-1/2} \mathbf{X}'_k \Sigma^{-1}(q) \mathbf{d}_k \\
&= \sigma^{-2}(q) A_q(1)^2 D_T^{-1/2} \begin{pmatrix} \text{sgn}(k - k_0) d_{\mu,0} \delta_k v_T + \frac{1}{2} d_{\beta,0} \delta_k^2 v_T^3 \\ \text{sgn}(k - k_0) \frac{1}{2} d_{\mu,0} \delta_k (k + k_0) v_T + \frac{1}{6} d_{\beta,0} \delta_k^2 (2k + k_0) v_T^3 \\ \mathbf{1}_{\{k < k_0\}} \left(-d_{\mu,0} \delta_k v_T + \frac{1}{2} d_{\beta,0} \delta_k^2 v_T^3 \right) \\ \mathbf{1}_{\{k < k_0\}} \left(-\frac{1}{2} d_{\mu,0} \delta_k^2 v_T + \frac{1}{6} d_{\beta,0} \delta_k^3 v_T^3 \right) \end{pmatrix} + \mathbf{R}(\delta_k, v_T),
\end{aligned} \tag{2.59}$$

where under $d_{\beta,0} \neq 0$,

$$\begin{aligned} \sup_{\{\delta_k: \delta_k > M_0 v_T^{-2}\}} \left\| T^{1/2} (\delta_k v_T^3)^{-1} \mathbf{R}(\delta_k, v_T) \right\| &\leq C, \\ \sup_{\{\delta_k: \delta_k \leq M_0 v_T^{-2}\}} \left\| \mathbf{R}(\delta_k, v_T) \right\| &\leq C (\sqrt{T} v_T)^{-1}, \end{aligned} \quad (2.60)$$

and as a result,

$$\begin{aligned} \sup_{\{\delta_k: \delta_k > M_0 v_T^{-2}\}} \left\| T^{1/2} (\delta_k^2 v_T^3)^{-1} \mathbf{D}_T^{-1/2} \mathbf{X}'_k \Sigma^{-1}(q) \mathbf{d}_k \right\| &\leq C, \\ \sup_{\{\delta_k: \delta_k \leq M_0 v_T^{-2}\}} \left\| (\sqrt{T} v_T) \mathbf{D}_T^{-1/2} \mathbf{X}'_k \Sigma^{-1}(q) \mathbf{d}_k \right\| &\leq C. \end{aligned} \quad (2.61)$$

If $d_{\beta,0} = 0$, the results remain unchanged for the case $\{\delta_k : \delta_k \leq M_0 v_T^{-2}\}$, but

$$\begin{aligned} \sup_{\{\delta_k: \delta_k > M_0 v_T^{-2}\}} \left\| T^{1/2} v_T^{-1} \mathbf{R}(\delta_k, v_T) \right\| &\leq C, \\ \sup_{\{\delta_k: \delta_k > M_0 v_T^{-2}\}} \left\| T^{1/2} (\delta_k v_T)^{-1} \mathbf{D}_T^{-1/2} \mathbf{X}'_k \Sigma^{-1}(q) \mathbf{d}_k \right\| &\leq C. \end{aligned}$$

Proof We only show the case $d_{\beta,0} \neq 0$. Recall $d_{k,t}$ in (2.50). We consider $\mathbf{D}_T^{-1/2} \mathbf{X}'_k \Sigma^{-1}(q) \mathbf{d}_k$ element-wise. If $k > k_0$, we have $d_{k,t} = [\mu_{b,0} + \beta_{b,0}(t - k_0)] (\mathbf{1}_{\{t > k_0\}} - \mathbf{1}_{\{t > k\}})$, and then using Lemma 2.2,

$$\begin{aligned} A_q(L) d_{k,t} &= A_q(1) d_{k,t} + \mu_{b,0} \tilde{A}_q(L) (\mathbf{1}_{\{t=k_0+1\}} - \mathbf{1}_{\{t=k+1\}}) \\ &\quad + \beta_{b,0} \tilde{A}_q(L) \mathbf{1}_{\{k_0+1 \leq t \leq k\}} - \beta_{b,0} \delta_k \tilde{A}_q(L) \mathbf{1}_{\{t=k+1\}} =: A_q(1) (d_{k,t} + \tilde{d}_{k,t}), \end{aligned} \quad (2.62)$$

where

$$\begin{aligned} \tilde{d}_{k,t} &= A_q(1)^{-1} \left[\mu_{b,0} \tilde{A}_q(L) (\mathbf{1}_{\{t=k_0+1\}} - \mathbf{1}_{\{t=k+1\}}) + \beta_{b,0} \tilde{A}_q(L) \mathbf{1}_{\{k_0+1 \leq t \leq k\}} - \beta_{b,0} \delta_k \tilde{A}_q(L) \mathbf{1}_{\{t=k+1\}} \right] \\ &= A_q(1)^{-1} \left[\mu_{b,0} \tilde{A}_q(L) \mathbf{1}_{\{t=k_0+1\}} - (\mu_{b,0} + \beta_{b,0} \delta_k) \tilde{A}_q(L) \mathbf{1}_{\{t=k+1\}} + \beta_{b,0} \tilde{A}_q(L) \mathbf{1}_{\{k_0+1 \leq t \leq k\}} \right]. \end{aligned} \quad (2.63)$$

For any $k > k_0$ and $k \in \Gamma_T$, since $A_q(L) d_{k,t} = 0$ for all $t \leq k_0$, by Lemmas 2.2 and 2.6,

we have

$$\begin{aligned}
D_T^{-1/2} X_k' \Sigma^{-1}(q) d_k &= \sigma^{-2}(q) D_T^{-1/2} \sum_{t=k_0+1}^T [A_q(L) x_{k,t}] [A_q(L) d_{k,t}] \\
&= \sigma^{-2}(q) D_T^{-1/2} \sum_{t=k_0+1}^T \left\{ A_q(1)^2 \begin{pmatrix} d_{k,t} \\ t d_{k,t} \\ \mathbf{1}_{\{t>k\}} d_{k,t} \\ (t-k) \mathbf{1}_{\{t>k\}} d_{k,t} \end{pmatrix} + \tilde{\mathbf{R}}_t \right\} \\
&= \sigma^{-2}(q) A_q(1)^2 D_T^{-1/2} \begin{pmatrix} d_{\mu,0} \delta_k v_T + \frac{1}{2} d_{\beta,0} \delta_k^2 v_T^3 \\ \frac{1}{2} d_{\mu,0} \delta_k (k+k_0) v_T + \frac{1}{6} d_{\beta,0} \delta_k^2 (2k+k_0) v_T^3 \\ \mathbf{0}_{2 \times 1} \end{pmatrix} + \mathbf{R}(\delta_k, v_T),
\end{aligned} \tag{2.64}$$

where

$$\tilde{\mathbf{R}}_t = \begin{pmatrix} A_q(1)^2 \tilde{d}_{k,t} \\ A_q(1)^2 t \tilde{d}_{k,t} + A_q(1) \tilde{A}_q(1) (d_{k,t} + \tilde{d}_{k,t}) \\ A_q(1)^2 \mathbf{1}_{\{t>k\}} \tilde{d}_{k,t} + A_q(1) [\tilde{A}_q(L) \mathbf{1}_{\{t=k+1\}}] (d_{k,t} + \tilde{d}_{k,t}) \\ A_q(1)^2 (t-k) \mathbf{1}_{\{t>k\}} \tilde{d}_{k,t} + A_q(1) [\tilde{A}_q(L) \mathbf{1}_{\{t>k\}}] (d_{k,t} + \tilde{d}_{k,t}) \end{pmatrix}, \tag{2.65}$$

$$\mathbf{R}(\delta_k, v_T) = \underbrace{\sigma^{-2}(q) A_q(1)^2 D_T^{-1/2} \begin{pmatrix} R_p(\delta_k, v_T) \\ R_w(\delta_k, v_T) \\ \mathbf{0}_{2 \times 1} \end{pmatrix}}_{\mathbf{R}_1(\delta_k, v_T)} + \underbrace{\sigma^{-2}(q) D_T^{-1/2} \sum_{t=k_0+1}^T \tilde{\mathbf{R}}_t}_{\mathbf{R}_2(\delta_k, v_T)}, \quad k > k_0, \tag{2.66}$$

with R_p and R_w that are given in Lemma 2.6. For (2.60), we note that the supremum of $\|\mathbf{R}(\delta_k, v_T)\|$ is bounded above by the sum (up to some multiplicative positive constant) of the respective suprema of $\|\mathbf{R}_1(\delta_k, v_T)\|$ and $\|\mathbf{R}_2(\delta_k, v_T)\|$. By looking at $\mathbf{R}_1(\delta_k, v_T)$ element-wise, the supremum of each element is nothing but the results given in Equation (2.58) in Lemma 2.6.

It remains to consider the second component $\mathbf{R}_2(\delta_k, v_T)$ in (2.66) which contains four elements by (2.65). It is sufficient to verify each of these elements can be bounded similarly as in (2.60). By the inequality in Lemma 2.2 (iii), the first element

$A_q(1)^2 T^{-1/2} \sum_{t=k_0+1}^T \tilde{d}_{k,t} \leq CT^{-1/2} \sum_{t=k_0+1}^T |\tilde{d}_{k,t}|$ and is further bounded by

$$\begin{aligned}
 & CT^{-1/2} \sum_{t=k_0+1}^T \left\{ |\mu_{b,0}| |\tilde{A}_q(L) \mathbf{1}_{\{t=k_0+1\}}| + |\mu_{b,0} + \beta_{b,0} \delta_k| |\tilde{A}_q(L) \mathbf{1}_{\{t=k+1\}}| + |\beta_{b,0}| |\tilde{A}_q(L) \mathbf{1}_{\{k_0+1 \leq t \leq k\}}| \right\} \\
 & \leq CT^{-1/2} \sum_{t=k_0+1}^T \sum_{j=1}^q |\tilde{a}_j(q)| \left\{ |\mu_{b,0}| \mathbf{1}_{\{t=k_0+j\}} + |\mu_{b,0} + \beta_{b,0} \delta_k| \mathbf{1}_{\{t=k+j\}} + |\beta_{b,0}| \mathbf{1}_{\{k_0+j \leq t \leq k+j-1\}} \right\} \\
 & = CT^{-1/2} \sum_{j=1}^q |\tilde{a}_j(q)| \left(|\mu_{b,0}| + |\mu_{b,0} + \beta_{b,0} \delta_k| + |\beta_{b,0} \delta_k| \right) \\
 & \leq CT^{-1/2} \left(|\mu_{b,0}| + |\beta_{b,0} \delta_k| \right).
 \end{aligned}$$

Because

$$\begin{aligned}
 & \sup_{\{\delta_k: \delta_k > M_0 v_T^{-2}\}} \left| T^{1/2} (\delta_k v_T^3)^{-1} T^{-1/2} (|\mu_{b,0}| + |\beta_{b,0} \delta_k|) \right| \leq C, \\
 & \sup_{\{\delta_k: \delta_k \leq M_0 v_T^{-2}\}} \left| T^{-1/2} (|\mu_{b,0}| + |\beta_{b,0} \delta_k|) \right| \leq C v_T / T^{1/2} \leq C (\sqrt{T} v_T)^{-1},
 \end{aligned} \tag{2.67}$$

we can see this element $A_q(1)^2 T^{-1/2} \sum_{t=k_0+1}^T \tilde{d}_{k,t}$ indeed satisfies (2.60). Similarly, by Lemma 2.6, the second element of $\mathbf{R}_2(\delta_k, v_T)$ is bounded by

$$\begin{aligned}
 & CT^{-3/2} \left(\sum_{t=k_0+1}^T |\tilde{d}_{k,t}| + \sum_{t=k_0+1}^T |d_{k,t}| \right) \\
 & \leq CT^{-1/2} (|d_{\mu,0}| v_T + |d_{\beta,0}| \delta_k v_T^3) + CT^{-3/2} (|d_{\mu,0}| \delta_k v_T + |d_{\beta,0}| \delta_k^2 v_T^3 + |R_p(\delta_k, v_T)|).
 \end{aligned}$$

Then similar results of (2.67) can be concluded for this element. For the third element of $\mathbf{R}_2(\delta_k, v_T)$, we have an upper bound

$$C \left(\left| T^{-1/2} \sum_{t=k_0+1}^T \mathbf{1}_{\{t>k\}} \tilde{d}_{k,t} \right| + \left| T^{-1/2} \sum_{t=k_0+1}^T [\tilde{A}_q(L) \mathbf{1}_{\{t=k+1\}}] d_{k,t} \right| + \left| T^{-1/2} \sum_{t=k_0+1}^T [\tilde{A}_q(L) \mathbf{1}_{\{t=k+1\}}] \tilde{d}_{k,t} \right| \right). \tag{2.68}$$

By the definition of $d_{k,t}$ and the fact that $\tilde{A}_q(L) \mathbf{1}_{\{t=k+1\}} = 0$ for $t \leq k$, we find the middle term in the brackets

$$\left| T^{-1/2} \sum_{t=k_0+1}^T [\tilde{A}_q(L) \mathbf{1}_{\{t=k+1\}}] d_{k,t} \right| = \left| T^{-1/2} \sum_{t=k_0+1}^k [\tilde{A}_q(L) \mathbf{1}_{\{t=k+1\}}] d_{k,t} \right| = 0.$$

Together with Lemma 2.2 (iii), it is not hard to find (2.68) is eventually bounded by

$T^{-1/2} \sum_{t=k_0+1}^T |\tilde{d}_{k,t}|$ which has the upper bounds given in (2.67), leading to the same bound in (2.60). Finally, the last element of $\mathbf{R}_2(\delta_k, v_T)$ is bounded by

$$\begin{aligned} & C \left(\left| T^{-1/2} \sum_{t=k_0+1}^T \frac{t-k}{T} \mathbf{1}_{\{t>k\}} \tilde{d}_{k,t} \right| + \left| T^{-3/2} \sum_{t=k_0+1}^T [\tilde{A}_q(L) \mathbf{1}_{\{t>k\}}] d_{k,t} \right| + \left| T^{-3/2} \sum_{t=k_0+1}^T [\tilde{A}_q(L) \mathbf{1}_{\{t>k\}}] \tilde{d}_{k,t} \right| \right) \\ & \leq C \left(T^{-1/2} \sum_{t=k_0+1}^T \left| \frac{t-k}{T} \mathbf{1}_{\{t>k\}} \tilde{d}_{k,t} \right| + \left| T^{-3/2} \sum_{t=k_0+1}^k [\tilde{A}_q(L) \mathbf{1}_{\{t>k\}}] d_{k,t} \right| + T^{-3/2} \sum_{t=k_0+1}^T \left| [\tilde{A}_q(L) \mathbf{1}_{\{t>k\}}] \tilde{d}_{k,t} \right| \right) \\ & \leq C \left\{ T^{-1/2} \sum_{t=k_0+1}^T |\tilde{d}_{k,t}| + 0 + \left(T^{-1} \sum_{t=k_0+1}^T |\tilde{A}_q(L) \mathbf{1}_{\{t>k\}}| \right) \left(T^{-1/2} \sum_{t=k_0+1}^T |\tilde{d}_{k,t}| \right) \right\} \leq CT^{-1/2} \sum_{t=k_0+1}^T |\tilde{d}_{k,t}|, \end{aligned}$$

where we use $T^{-1} \sum_{t=k_0+1}^T |\tilde{A}_q(L) \mathbf{1}_{\{t>k\}}| \leq C$ (by Lemma 2.2 (iv)) to obtain the final inequality. Now, (2.60) should be clearly seen using (2.67).

For $k < k_0$, summations should be taken from $k+1$ (instead of k_0+1) to T . The results above hold with a slight modification. We thus omit the proof. Finally, (2.61) is a direct result of (2.59) and (2.60). ■

Lemma 2.8

Given Assumptions 2.1 to 2.4 hold, for any $\epsilon > 0$, there exists a positive constant C_ϵ such that

$$\mathbb{P} \left(\sup_{\{\delta_k: \delta_k > M_0 v_T^{-2}\}} |(\delta_k^2 v_T^4)^{-1} \mathbf{d}'_k \Sigma^{-1}(q) \epsilon| \geq C_\epsilon \right) < \epsilon, \quad \text{if } d_{\beta,0} \neq 0. \quad (2.69)$$

$$\text{If } d_{\beta,0} = 0, \text{ then } \mathbb{P} \left(\sup_{\{\delta_k: \delta_k > M_0 v_T^{-2}\}} |(\delta_k v_T^2)^{-1} \mathbf{d}'_k \Sigma^{-1}(q) \epsilon| \geq C_\epsilon \right) < \epsilon.$$

Proof We only show the case where $k > k_0$ and $d_{\beta,0} \neq 0$. The cases where $k < k_0$ or $d_{\beta,0} = 0$ are similar. For convenience, let $\eta_t^* = A_q(L) \epsilon_t$ be the filtered errors. We have

$$\begin{aligned} \mathbf{d}'_k \Sigma^{-1}(q) \epsilon &= \sigma^{-2}(q) \sum_{t=k_0+1}^T [A_q(L) d_{k,t}] \eta_t^* \\ &= \sigma^{-2}(q) A_q(1) \sum_{t=k_0+1}^T d_{k,t} \eta_t^* + \sigma^{-2}(q) \sum_{t=k_0+1}^T [A_q(1) \tilde{d}_{k,t}] \eta_t^*, \end{aligned}$$

where $A_q(L)d_{k,t}$ and $\tilde{d}_{k,t}$ are given in (2.62) and (2.63). To show (2.69), it suffices to prove

$$\sup_{\{\delta_k: \delta_k > M_0 v_T^{-2}\}} \left| (\delta_k^2 v_T^4)^{-1} \sum_{t=k_0+1}^T d_{k,t} \eta_t^* \right| = O_p(1), \quad (2.70)$$

$$\sup_{\{\delta_k: \delta_k > M_0 v_T^{-2}\}} \left| (\delta_k^2 v_T^4)^{-1} \sum_{t=k_0+1}^T [A_q(1)\tilde{d}_{k,t}] \eta_t^* \right| = O_p(1). \quad (2.71)$$

Recall $d_{k,t} = [\mu_{b,0} + \beta_{b,0}(t - k_0)] \mathbf{1}_{\{k_0+1 \leq t \leq k\}}$ from (2.50). Using the identity $k = k_0 + \delta_k$, the LHS of (2.70) is bounded by

$$\frac{|d_{\mu,0}|}{M_0 v_T} \sup_{\{\delta_k: \delta_k > M_0 v_T^{-2}\}} \left| \delta_k^{-1} \sum_{t=k_0+1}^{k_0+\delta_k} \eta_t^* \right| + \frac{|d_{\beta,0}|}{v_T} \sup_{\{\delta_k: \delta_k > M_0 v_T^{-2}\}} \left| \delta_k^{-1} \sum_{t=k_0+1}^{k_0+\delta_k} \frac{t - k_0}{\delta_k} \eta_t^* \right|.$$

Along with the literature, such as the proof of Proposition 4(ii) in Bai and Perron (1998), we will employ a Hájek and Rényi (HR) type of inequality for the sums of filtered error process $\{\eta_t^*\}$ and the trend-scaled sequence $\left\{\frac{t-k_0}{\delta_k} \eta_t^*\right\}$. An appropriate HR inequality to be used here has been established in Proposition 1 of Bai (1994), which will be used repeatedly in the remaining proofs. Three remarks about the validity of applying the inequality are in place: first, note that $\eta_t^* = A_q(L)A(L)^{-1}\eta_t = B_q(L)\eta_t$, where $B_q(L)$ is defined in Lemma 2.1. By Lemma 2.1 (ii), $\{\eta_t^*\}$ is again a linear process of m.d.s., with $\sum_{j=1}^{\infty} j |b_j(q)| < \infty$ satisfying Assumption (B) of Bai (1994); second, although the inequality is stated for linear processes of *i.i.d.* sequence, it also holds for m.d.s. with finite second moments; third, a direct application of the inequality to the process $\left\{\frac{t-k_0}{\delta_k} \eta_t^*\right\}$ is not appropriate. However, with some modifications of the proof of Bai (1994), we can show the HR inequality again holds for the sequence $\left\{\frac{t-k_0}{\delta_k} \eta_t^*\right\}$ (see Lemma 2.12 in Appendix Section 2.E). As a result, for any $\epsilon > 0$, there exist $\alpha_\epsilon > 0$ and $C_1, C_2 < \infty$ such that

$$\begin{aligned} & \mathbb{P} \left(\sup_{\{\delta_k: \delta_k > M_0 v_T^{-2}\}} \left| (\delta_k^2 v_T^4)^{-1} \sum_{t=k_0+1}^T d_{k,t} \eta_t^* \right| > \alpha_\epsilon \right) \\ & \leq \mathbb{P} \left(\sup_{\{\delta_k: \delta_k > M_0 v_T^{-2}\}} \left| \delta_k^{-1} \sum_{t=k_0+1}^{k_0+\delta_k} \eta_t^* \right| > \frac{M_0 \alpha_\epsilon v_T}{2|d_{\mu,0}|} \right) + \mathbb{P} \left(\sup_{\{\delta_k: \delta_k > M_0 v_T^{-2}\}} \left| \delta_k^{-1} \sum_{t=k_0+1}^{k_0+\delta_k} \frac{t - k_0}{\delta_k} \eta_t^* \right| > \frac{\alpha_\epsilon v_T}{2|d_{\beta,0}|} \right) \\ & \leq C_1 \frac{4d_{\mu,0}^2}{M_0^2 \alpha_\epsilon^2 v_T^2} \frac{1}{M_0 v_T^{-2}} + C_2 \frac{4d_{\beta,0}^2}{\alpha_\epsilon^2 v_T^2} \frac{1}{M_0 v_T^{-2}} \leq \frac{C}{\alpha_\epsilon^2} < \epsilon, \end{aligned}$$

which proves (2.70). Moreover, by the definition of $\widetilde{d}_{k,t}$, we can have

$$\begin{aligned}
& \sum_{t=k_0+1}^T [A_q(1)\widetilde{d}_{k,t}]\eta_t^* \\
&= \sum_{t=k_0+1}^T \left[\mu_{b,0}\widetilde{A}_q(L)\mathbf{1}_{\{t=k_0+1\}} - (\mu_{b,0} + \beta_{b,0}\delta_k)\widetilde{A}_q(L)\mathbf{1}_{\{t=k+1\}} + \beta_{b,0}\widetilde{A}_q(L)\mathbf{1}_{\{k_0+1 \leq t \leq k\}} \right] \eta_t^* \\
&= \sum_{t=k_0+1}^T \sum_{j=1}^q \widetilde{a}_j(q) \left[\mu_{b,0}\mathbf{1}_{\{t=k_0+j\}} - (\mu_{b,0} + \beta_{b,0}\delta_k)\mathbf{1}_{\{t=k+j\}} + \beta_{b,0}\mathbf{1}_{\{k_0+j \leq t \leq k+j-1\}} \right] \eta_t^* \\
&= \sum_{j=1}^q \widetilde{a}_j(q) \left[\mu_{b,0}\eta_{k_0+j}^* - (\mu_{b,0} + \beta_{b,0}\delta_k)\eta_{k+j}^* + \beta_{b,0} \sum_{t=k_0+j}^{k+j-1} \eta_t^* \right] \quad (2.72)
\end{aligned}$$

$$= \mu_{b,0} \sum_{j=1}^q \widetilde{a}_j(q)\eta_{k_0+j}^* - (\mu_{b,0} + \beta_{b,0}\delta_k) \sum_{j=1}^q \widetilde{a}_j(q)\eta_{k+j}^* + \beta_{b,0} \sum_{j=1}^q \widetilde{a}_j(q) \sum_{t=k_0+j}^{k+j-1} \eta_t^*, \quad (2.73)$$

where (2.72) is obtained by interchanging the order of summation. To establish (2.71), three primary components shall be considered according to (2.73). First, by $\delta_{k_0+j} = j$,

$$\begin{aligned}
& \sup_{\{\delta_k: \delta_k > M_0 v_T^{-2}\}} \left| (\delta_k^2 v_T^4)^{-1} \mu_{b,0} \sum_{j=1}^q \widetilde{a}_j(q)\eta_{k_0+j}^* \right| \leq C v_T \sum_{j=1}^q |\widetilde{a}_j(q)| |\eta_{k_0+j}^*| \\
&= C v_T \sum_{j=1}^q j |\widetilde{a}_j(q)| |\delta_{k_0+j}^{-1} \eta_{k_0+j}^*| \\
&\leq C v_T \left(\sum_{j=1}^q j |\widetilde{a}_j(q)| \right) \left(\sup_{k \geq k_0+1} |\delta_k^{-1} \eta_k^*| \right) = O_p(v_T), \quad (2.74)
\end{aligned}$$

where by the HR inequality, we have

$$\sup_{k \geq k_0+1} |\delta_k^{-1} \eta_k^*| \leq \sup_{k \geq k_0+1} \left(\delta_k^{-1} \left| \sum_{j=k_0+1}^k \eta_j^* \right| + \delta_k^{-1} \left| \sum_{j=k_0}^{k-1} \eta_j^* \right| + \delta_k^{-1} \eta_{k_0}^* \right) = O_p(1). \quad (2.75)$$

For the second component in (2.73), by $\delta_{k+j} = |k+j-k_0| \leq \delta_k + j \leq 2j\delta_k$ if $j \geq 1$,

$$\begin{aligned}
 & \sup_{\{\delta_k: \delta_k > M_0 v_T^{-2}\}} \left| (\delta_k^2 v_T^4)^{-1} (\mu_{b,0} + \beta_{b,0} \delta_k) \sum_{j=1}^q \tilde{a}_j(q) \eta_{k+j}^* \right| \\
 & \leq C \sup_{\{\delta_k: \delta_k > M_0 v_T^{-2}\}} \left((\delta_k v_T)^{-1} \sum_{j=1}^q |\tilde{a}_j(q)| |\eta_{k+j}^*| \right) \\
 & = (|d_{\mu,0}| M_0^{-1} + |d_{\beta,0}|) \frac{1}{\delta_k v_T} \sum_{j=1}^q \left[|\tilde{a}_j(q)| \left| \frac{\epsilon_{k+j}}{\delta_{k+j}} \right| \delta_{k+j} \right] \\
 & \leq C \sup_{\{\delta_k: \delta_k > M_0 v_T^{-2}\}} \left((\delta_k v_T)^{-1} \sum_{j=1}^q \left[|\tilde{a}_j(q)| |\delta_{k+j}^{-1} \eta_{k+j}^*| (\delta_k + j) \right] \right) \\
 & = C \left[v_T^{-1} \sum_{j=1}^q |\tilde{a}_j(q)| |\delta_{k+j}^{-1} \epsilon_{k+j}| + (\delta_k v_T)^{-1} \sum_{j=1}^q j |\tilde{a}_j(q)| |\delta_{k+j}^{-1} \epsilon_{k+j}| \right] \\
 & \leq 2C \sup_{\{\delta_k: \delta_k > M_0 v_T^{-2}\}} \left(v_T^{-1} \sum_{j=1}^q j |\tilde{a}_j(q)| |\delta_{k+j}^{-1} \eta_{k+j}^*| \right) \\
 & \leq \left(2C \sum_{j=1}^q j |\tilde{a}_j(q)| \right) \left(v_T^{-1} \sup_{k > k_0 + M_0 v_T^{-2}} |\delta_k^{-1} \eta_k^*| \right) = O_p(1), \tag{2.76}
 \end{aligned}$$

where $v_T^{-1} \sup_{k > k_0 + M_0 v_T^{-2}} |\delta_k^{-1} \eta_k^*| = O_p(1)$ similarly as (2.75) by the HR inequality. Finally,

$$\begin{aligned}
 & \sup_{\{\delta_k: \delta_k > M_0 v_T^{-2}\}} \left| (\delta_k^2 v_T^4)^{-1} \beta_{b,0} \sum_{j=1}^q \tilde{a}_j(q) \sum_{t=k_0+j}^{k+j-1} \eta_t^* \right| \leq C v_T \sup_{\{\delta_k: \delta_k > M_0 v_T^{-2}\}} \sup_{1 \leq j \leq q} \left(\delta_k^{-1} \left| \sum_{t=k_0+j}^{k+j-1} \eta_t^* \right| \right) \\
 & = C v_T^2 \sup_{1 \leq j \leq q} \sup_{\{\delta_k: \delta_k > M_0 v_T^{-2}\}} \left(v_T^{-1} \delta_k^{-1} \left| \sum_{s=1}^{\delta_k} \eta_{k_0+j-1+s}^* \right| \right) = O_p(v_T^2) \tag{2.77}
 \end{aligned}$$

where, using the HR inequality, $\sup_{\{\delta_k: \delta_k > M_0 v_T^{-2}\}} \left(v_T^{-1} \delta_k^{-1} \left| \sum_{s=1}^{\delta_k} \eta_{k_0+j-1+s}^* \right| \right)$ can be bounded with probability 1 by a constant that does not depend on the index j . Combining (2.74) – (2.77), we have (2.71) and thus the lemma. \blacksquare

Lemma 2.9

Suppose Assumptions 2.1 to 2.4 hold. For any $k \in \Gamma_T$,

$$D_T^{-1/2} \mathbf{X}'_k \Sigma^{-1}(q) \boldsymbol{\varepsilon} = \sigma^{-2}(q) \begin{pmatrix} \frac{1}{T^{1/2}} \sum_{t=1}^T [A_q(L) \cdot 1][A_q(L) \epsilon_t] \\ \frac{1}{T^{3/2}} \sum_{t=1}^T [A_q(L) \cdot t][A_q(L) \epsilon_t] \\ \frac{1}{T^{1/2}} \sum_{t=k+1}^T [A_q(L) \cdot 1][A_q(L) \epsilon_t] \\ \frac{1}{T^{3/2}} \sum_{t=k+1}^T [A_q(L)(t-k)][A_q(L) \epsilon_t] \end{pmatrix} + \sigma^{-2}(q) \widetilde{\mathbf{R}}_k^*, \quad (2.78)$$

where $\widetilde{\mathbf{R}}_k^*$ is defined in (2.82) below.

(i) The remainder term $\widetilde{\mathbf{R}}_k^*$ has asymptotic orders:

$$\sup_{k \in V_T(M_0)} \widetilde{\mathbf{R}}_k^* = O_p(\sqrt{\ln T}), \quad \sup_{k \in K_T(M_0)} \widetilde{\mathbf{R}}_k^* = o_p(1). \quad (2.79)$$

(ii) The first term of (2.78) is $O_p(1)$ uniform over Γ_T . Namely, by (2.79), $D_T^{-1/2} \mathbf{X}'_k \Sigma^{-1}(q) \boldsymbol{\varepsilon} = O_p(\sqrt{\ln T})$ uniformly over $V_T(M_0)$, and $D_T^{-1/2} \mathbf{X}'_k \Sigma^{-1}(q) \boldsymbol{\varepsilon} = O_p(1)$ uniformly over $K_T(M_0)$.

(iii) $D_T^{-1/2} \mathbf{X}'_k \Sigma^{-1}(q) \boldsymbol{\varepsilon}$ and $D_T^{-1/2} \mathbf{X}'_{k_0} \Sigma^{-1}(q) \boldsymbol{\varepsilon}$ are asymptotically equivalent for any $k \in K_T(M_0)$. More specifically,

$$D_T^{-1/2} \mathbf{X}'_{k_0} \Sigma^{-1}(q) \boldsymbol{\varepsilon} - D_T^{-1/2} \mathbf{X}'_k \Sigma^{-1}(q) \boldsymbol{\varepsilon} = o_p(1), \quad (2.80)$$

where $o_p(1)$ is uniform over $K_T(M_0)$.

Proof Recall that $\eta_t^* = A_q(L) \epsilon_t$. Based on the same reasoning as the proof of Lemma

2.5, we assume that $A_q(L)$ and $\sigma^{-2}(q)$ applies for all $t = 1, 2, \dots$. For any $k \in \Gamma_T$,

$$\begin{aligned} D_T^{-1/2} \mathbf{X}'_k \Sigma^{-1}(q) \varepsilon &= \sigma^{-2}(q) D_T^{-1/2} \sum_{t=1}^T [A_q(L) \mathbf{x}_{k,t}] \eta_t^* \\ &= \sigma^{-2}(q) \begin{pmatrix} \frac{1}{T^{1/2}} \sum_{t=1}^T [A_q(L) \cdot 1] \eta_t^* \\ \frac{1}{T^{3/2}} \sum_{t=1}^T [A_q(L) \cdot t] \eta_t^* \\ \frac{1}{T^{1/2}} \sum_{t=1}^T [A_q(L) \mathbf{1}_{\{t>k\}}] \eta_t^* \\ \frac{1}{T^{3/2}} \sum_{t=1}^T [A_q(L)(t-k) \mathbf{1}_{\{t>k\}}] \eta_t^* \end{pmatrix}. \end{aligned} \quad (2.81)$$

Note that the first two entries of (2.81) are equal to the ones in the vector of (2.78).

Now we apply Lemma 2.2 (i) and (iii) to the third entry and obtain

$$\frac{1}{T^{1/2}} \sum_{t=1}^T [A_q(L) \mathbf{1}_{\{t>k\}}] \eta_t^* = \frac{1}{T^{1/2}} \sum_{t=k+1}^T [A_q(L) \cdot 1] \eta_t^* + \frac{1}{T^{1/2}} \sum_{t=1}^T [\tilde{A}_q(L) \mathbf{1}_{\{t=k+1\}}] \eta_t^*.$$

Moreover, using Lemma 2.2 (iv) and the property $A_q(1)(t-k) = A_q(L)(t-k) - \tilde{A}_q(1)$, we can rewrite the last entry as

$$\begin{aligned} &\frac{1}{T^{3/2}} \sum_{t=1}^T [A_q(L)(t-k) \mathbf{1}_{\{t>k\}}] \eta_t^* \\ &= \frac{1}{T^{3/2}} \sum_{t=k+1}^T [A_q(L)(t-k)] \eta_t^* - \tilde{A}_q(1) \frac{1}{T^{3/2}} \sum_{t=k+1}^T \eta_t^* + \frac{1}{T^{3/2}} \sum_{t=1}^T [\tilde{A}_q(L) \mathbf{1}_{\{t>k\}}] \eta_t^*. \end{aligned}$$

Plugging these two additional results into (2.81) leads to (2.78), where

$$\tilde{\mathbf{R}}_k^* = \begin{pmatrix} \mathbf{0}_{2 \times 1} \\ \frac{1}{T^{1/2}} \sum_{t=1}^T [\tilde{A}_q(L) \mathbf{1}_{\{t=k+1\}}] \eta_t^* \\ -\tilde{A}_q(1) \frac{1}{T^{3/2}} \sum_{t=k+1}^T \eta_t^* + \frac{1}{T^{3/2}} \sum_{t=1}^T [\tilde{A}_q(L) \mathbf{1}_{\{t>k\}}] \eta_t^* \end{pmatrix}. \quad (2.82)$$

Now we move on to prove (2.79). Without loss of generality, we only show the case $k > k_0$. By norm equivalence, we consider the elements in (2.82) individually. Using the identity $\tilde{A}_q(L) \mathbf{1}_{\{t=k+1\}} = \sum_{j=1}^q \tilde{a}_j(q) \mathbf{1}_{\{t=k+j\}}$, and then interchanging the order of

summation, we have an upper bound for the third element of $\tilde{\mathbf{R}}_k^*$ as

$$\begin{aligned}
\left| \frac{1}{T^{1/2}} \sum_{t=1}^T [\tilde{A}_q(L) \mathbf{1}_{\{t=k+1\}}] \eta_t^* \right| &\leq \frac{1}{T^{1/2}} \sum_{t=1}^T \left[\sum_{j=1}^q |\tilde{a}_j(q)| \mathbf{1}_{\{t=k+j\}} \right] |\eta_t^*| \\
&= \frac{1}{T^{1/2}} \sum_{j=1}^q |\tilde{a}_j(q)| |\eta_{k+j}^*| \\
&\leq \frac{1}{T^{1/2}} \sum_{j=1}^q (\delta_k + j)^{1/2} |\tilde{a}_j(q)| \left| \delta_{k+j}^{-1/2} \eta_{k+j}^* \right| \\
&\leq 2 \sqrt{\frac{\delta_k}{T}} \left(\sum_{j=1}^q j^{1/2} |\tilde{a}_j(q)| \right) \left(\sup_{t \geq k_0+1} |\delta_t^{-1/2} \eta_t^*| \right) \\
&\leq C \sqrt{\frac{\delta_k}{T}} \left(\sup_{t \geq k_0+1} |\delta_t^{-1/2} \eta_t^*| \right), \tag{2.83}
\end{aligned}$$

where, by c_r -inequality, $\delta_{k+j}^{1/2} = |k+j-k_0|^{1/2} \leq (\delta_k + j)^{1/2} \leq \delta_k^{1/2} + j^{1/2} \leq 2\delta_k^{1/2} j^{1/2}$ if $j \geq 1$. Similarly as (2.75), by the HR inequality, we have $\sup_{t \geq k_0+1} |\delta_t^{-1/2} \eta_t^*| = O_p(\sqrt{\ln T})$. From (2.83),

$$\begin{aligned}
\sup_{k \in V_T(M_0)} \left| \frac{1}{T^{1/2}} \sum_{t=1}^T [\tilde{A}_q(L) \mathbf{1}_{\{t=k+1\}}] \eta_t^* \right| &= O_p(\sqrt{\ln T}), \\
\sup_{k \in K_T(M_0)} \left| \frac{1}{T^{1/2}} \sum_{t=1}^T [\tilde{A}_q(L) \mathbf{1}_{\{t=k+1\}}] \eta_t^* \right| &= O_p\left(\frac{\sqrt{\ln T}}{\sqrt{T} v_T}\right) = o_p(1). \tag{2.84}
\end{aligned}$$

Next, by triangular inequality,

$$\left| \frac{1}{T^{3/2}} \sum_{t=k+1}^T \eta_t^* \right| \leq \frac{k^{1/2}}{T^{3/2}} \left| \frac{1}{k^{1/2}} \sum_{t=1}^k \eta_t^* \right| + \frac{1}{T} \left| \frac{1}{T^{1/2}} \sum_{t=1}^T \eta_t^* \right|. \tag{2.85}$$

By the HR inequality, for any $\epsilon > 0$, there exists an $\alpha_\epsilon > 0$ such that $\mathbb{P}\left(\sup_{k \in \Gamma_T} \frac{1}{k^{1/2}} \left| \sum_{t=1}^k \eta_t^* \right| > \alpha_\epsilon\right) \leq C \frac{1}{\alpha_\epsilon^2} \sum_{k=[\delta_1 T]}^T \frac{1}{k} \leq C \delta_1 \frac{1}{\alpha_\epsilon^2} < \epsilon$, where the Harmonic number $\sum_{k=[\delta_1 T]}^T \frac{1}{k} \rightarrow \ln \delta_1^{-1}$ as $T \rightarrow \infty$ with δ_1 defined in Assumption 2.1. Moreover, Lemma 2.3 (i) implies $\left| \frac{1}{T^{1/2}} \sum_{t=1}^T \eta_t^* \right| = O_p(1)$. Together with (2.85), we have

$$\sup_{k \in \Gamma_T} \left| \frac{1}{T^{3/2}} \sum_{t=k+1}^T \eta_t^* \right| = O_p(T^{-1}). \tag{2.86}$$

Finally, by the identity $\widetilde{A}_q(L)\mathbf{1}_{\{t>k\}} = \sum_{j=1}^q \widetilde{a}_j(q)\mathbf{1}_{\{t>k+j-1\}}$ and (2.86),

$$\begin{aligned} \left| \frac{1}{T^{3/2}} \sum_{t=1}^T [\widetilde{A}_q(L)\mathbf{1}_{\{t>k\}}]\eta_t^* \right| &= \frac{1}{T^{3/2}} \left| \sum_{t=1}^T \sum_{j=1}^q [\widetilde{a}_j(q)\mathbf{1}_{\{t>k+j-1\}}]\eta_t^* \right| \\ &= \frac{1}{T^{3/2}} \left| \sum_{j=1}^q \widetilde{a}_j(q) \sum_{t=k+j}^T \eta_t^* \right| \\ &\leq C \sup_{k \in \Gamma_T} \left| \frac{1}{T^{3/2}} \sum_{t=k+1}^T \eta_t^* \right| = O_p(T^{-1}) \end{aligned} \quad (2.87)$$

uniformly over Γ_T . By (2.84), (2.86) and (2.87), we obtain (2.79).

For Part (ii), by Lemma 2.3 (i) and (ii), the first two components are $O_p(1)$ uniformly. The third entry $\sup_{k \in \Gamma_T} \left| \frac{1}{T^{1/2}} \sum_{t=k+1}^T [A_q(L) \cdot 1]\eta_t^* \right| = |A_q(1)| \sup_{k \in \Gamma_T} \left| \frac{1}{T^{1/2}} \sum_{t=k+1}^T \eta_t^* \right| = O_p(1)$ by (2.86). Similarly,

$$\begin{aligned} \sup_{k \in \Gamma_T} \left| \frac{1}{T^{3/2}} \sum_{t=k+1}^T [A_q(L)(t-k)]\eta_t^* \right| &\leq \sup_{k \in \Gamma_T} |A_q(1)| \left| \frac{1}{T^{3/2}} \sum_{t=k+1}^T (t-k)\eta_t^* \right| + |\widetilde{A}_q(1)| \sup_{k \in \Gamma_T} \left| \frac{1}{T^{3/2}} \sum_{t=k+1}^T \eta_t^* \right| \\ &= |A_q(1)| \sup_{k \in \Gamma_T} \left| \frac{1}{T^{3/2}} \sum_{t=k+1}^T (t-k)\eta_t^* \right| + O_p(T^{-1}). \end{aligned}$$

Apply the HR inequality for the trend-scaled sequence $\left\{ \frac{t-k}{T-k} \eta_t^* \right\}_{t=k+1}^T$ and subsequently obtain, $\forall \epsilon > 0$, there exists an $\alpha_\epsilon > 0$ such that

$$\begin{aligned} \mathbb{P} \left(\sup_{k \in \Gamma_T} \frac{1}{\sqrt{T-k}} \left| \sum_{t=k+1}^T \frac{t-k}{T-k} \eta_t^* \right| > \alpha_\epsilon \right) \\ \leq \mathbb{P} \left(\sup_{[\delta_2 T] \leq j \leq T} \frac{1}{\sqrt{j}} \left| \sum_{t=1}^{T-k} \frac{t}{j} \eta_{k+t}^* \right| > \alpha_\epsilon \right) \leq C \alpha_\epsilon^{-2} \sum_{t=[\delta_2 T]}^T t^{-1} < \epsilon, \end{aligned} \quad (2.88)$$

where δ_2 is given in Assumption 2.1. As a result, $\sup_{k \in \Gamma_T} \left| \frac{1}{T^{3/2}} \sum_{t=k+1}^T (t-k)\eta_t^* \right| = O_p(1)$. Combining these results, we conclude the second part.

We move on to show Part (iii). For $k > k_0$, a direct computation yields

$$\begin{aligned} D_T^{-1/2} \mathbf{X}'_{k_0} \boldsymbol{\Sigma}^{-1}(q) \boldsymbol{\varepsilon} - D_T^{-1/2} \mathbf{X}'_k \boldsymbol{\Sigma}^{-1}(q) \boldsymbol{\varepsilon} \\ = \sigma^{-2}(q) A_q(1) \begin{pmatrix} \mathbf{0}_{2 \times 1} \\ T^{-1/2} \sum_{t=k_0+1}^k \eta_t^* \\ T^{-3/2} \sum_{t=k_0+1}^k (t-k_0)\eta_t^* + T^{-3/2} [\delta_k + A_q(1)^{-1} \widetilde{A}_q(1)] \sum_{t=k_0+1}^k \eta_t^* \end{pmatrix} + o_p(1). \end{aligned}$$

By the HR inequality, we have $\sup_{k \in K_T(M_0)} \delta_k^{-1/2} \left| \sum_{t=k_0+1}^k \eta_t^* \right| = o_p(\sqrt{\ln T})$ and $\sup_{k \in K_T(M_0)} \delta_k^{-1/2} \left| \sum_{t=k_0+1}^k \frac{t-k_0}{\delta_k} \eta_t^* \right| = o_p(\sqrt{\ln T})$. We arrive at

$$D_T^{-1/2} X'_{k_0} \Sigma^{-1}(q) \varepsilon - D_T^{-1/2} X'_k \Sigma^{-1}(q) \varepsilon = o_p\left(\frac{\sqrt{\ln T}}{\sqrt{T} v_T}\right) = o_p(1) \quad (2.89)$$

uniformly over $K_T(M_0)$. The case $k < k_0$ is similar. \blacksquare

Lemma 2.10

Under Assumptions 2.1 to 2.4, for every M_0 , there exist $T^* > 0$ and $C_0 > 0$ such that

$$\inf_{k \in V_T(M_0)} \frac{(XX)_k}{\delta_k^3 v_T^6} \geq C_0 + O(M_0^{-1}) + O(M_0^{-2}), \quad \text{if } d_{\beta,0} \neq 0, \quad (2.90)$$

whenever $T > T^*$. If $d_{\beta,0} = 0$, then $\inf_{k \in V_T(M_0)} \frac{(XX)_k}{\delta_k v_T^2} \geq C_0$.

Proof We only prove the case of $d_{\beta,0} \neq 0$. It suffices to show that, for every $M_0 > 0$ and $k \in V_T(M_0)$, $(XX)_k / (\delta_k^3 v_T^6) \geq C_0 + O(M_0^{-1}) + O(M_0^{-2})$ for some $C_0 > 0$, where the terms $O(M_0^{-1})$ and $O(M_0^{-2})$ are uniform in k . Rewrite $(XX)_k$ as

$$\begin{aligned} (XX)_k &= d'_k \Sigma^{-1/2}(q) \left\{ I_T - \Sigma^{-1/2}(q) X_k \left[X'_k \Sigma^{-1}(q) X_k \right]^{-1} X'_k \Sigma^{-1/2}(q) \right\} \Sigma^{-1/2}(q) d_k \\ &= (d_k - X_k \beta_k)' \Sigma^{-1}(q) (d_k - X_k \beta_k) \end{aligned}$$

where $\beta_k = (\beta_{k,1}, \dots, \beta_{k,4})'$ is the coefficient vector of the orthogonal projection of $\Sigma^{-1/2}(q) d_k$ on the column space of $\Sigma^{-1/2}(q) X_k$, that is,

$$\begin{aligned} \beta_k &= \left[X'_k \Sigma^{-1}(q) X_k \right]^{-1} X'_k \Sigma^{-1}(q) d_k \\ &= D_T^{-1/2} \left[D_T^{-1/2} X'_k \Sigma^{-1}(q) X_k D_T^{-1/2} \right]^{-1} \left[D_T^{-1/2} X'_k \Sigma^{-1}(q) d_k \right]. \end{aligned}$$

Let $C_k = (\delta_k^3 v_T^6)^{-1} \|d_k - X_k \beta_k\|^2$. To continue, we first make a claim as follows:

For every $M_0 > 0$ and $k \in V_T(M_0)$, there exist $l_0, T^ > 0$ such that $C_k \geq l_0 + O(M_0^{-1}) + O(M_0^{-2})$ whenever $T > T^*$, where the terms $O(M_0^{-1})$ and $O(M_0^{-2})$ are uniform in k .*

Since $\mathbf{x}' \Sigma^{-1}(q) \mathbf{x} \geq \lambda_{\min}(\Sigma^{-1}(q)) \|\mathbf{x}\|^2$, under the claim above, we deduce

$$\begin{aligned} (XX)_k / (\delta_k^3 v_T^6) &\geq \lambda_{\min}(\Sigma^{-1}(q)) \|\mathbf{d}_k - \mathbf{X}_k \boldsymbol{\beta}_k\|^2 / (\delta_k^3 v_T^6) \\ &= \lambda_{\min}(\Sigma^{-1}(q)) C_k \geq \lambda_{\min}(\Sigma^{-1}(q)) l_0 \\ &=: C_0 + O(M_0^{-1}) + O(M_0^{-2}). \end{aligned}$$

It remains to derive the claim. We only show the case $k > k_0$. By calculation,

$$\begin{aligned} C_k &= (\delta_k^3 v_T^6)^{-1} \sum_{t=1}^T [d_{k,t} - \beta_{k,1} - \beta_{k,2}t - \beta_{k,3} \mathbf{1}_{\{t>k\}} - \beta_{k,4}(t-k) \mathbf{1}_{\{t>k\}}]^2 \\ &\geq (\delta_k^3 v_T^6)^{-1} \sum_{t=1}^k (d_{k,t} - \beta_{k,1} - \beta_{k,2}t)^2 \\ &= (\delta_k^3 v_T^6)^{-1} \left\{ \sum_{t=k_0+1}^k d_{k,t}^2 + \sum_{t=1}^k (\beta_{k,1} + \beta_{k,2}t)^2 - 2\beta_{k,1} \sum_{t=k_0+1}^k d_{k,t} - 2\beta_{k,2} \sum_{t=k_0+1}^k t d_{k,t} \right\} \\ &=: C_{k,1} + C_{k,2} - C_{k,3} - C_{k,4}. \end{aligned} \tag{2.91}$$

Lemma 2.6 implies

$$C_{k,1} = \frac{1}{3} d_{\beta,0}^2 + (\delta_k v_T^2)^{-1} d_{\mu,0} d_{\beta,0} + (\delta_k^2 v_T^4)^{-1} d_{\mu,0}^2 + O(v_T^2). \tag{2.92}$$

We consider to derive a lower bound of $C_{k,2} - C_{k,3} - C_{k,4}$. To achieve this, we begin with expressing $\beta_{k,1}$ and $\beta_{k,2}$ explicitly. Using Lemmas 2.5 and 2.7,

$$\begin{aligned} \beta_{k,1} &= \frac{4}{k} \left(d_{\mu,0} \delta_k v_T + \frac{1}{2} d_{\beta,0} \delta_k^2 v_T^3 \right) - \frac{6}{k^2} \left[\frac{1}{2} d_{\mu,0} \delta_k (k + k_0) v_T + \frac{1}{6} d_{\beta,0} \delta_k^2 (2k + k_0) v_T^3 \right] + R_1(\delta_k, v_T), \\ \beta_{k,2} &= -\frac{6}{k^2} \left(d_{\mu,0} \delta_k v_T + \frac{1}{2} d_{\beta,0} \delta_k^2 v_T^3 \right) + \frac{12}{k^3} \left[\frac{1}{2} d_{\mu,0} \delta_k (k + k_0) v_T + \frac{1}{6} d_{\beta,0} \delta_k^2 (2k + k_0) v_T^3 \right] + R_2(\delta_k, v_T). \end{aligned}$$

where the remainder terms are bounded as $\sup_{\{\delta_k: \delta_k > M_0 v_T^{-2}\}} |T(\delta_k v_T^3)^{-1} R_1(\delta_k, v_T)| \leq C$ and $\sup_{\{\delta_k: \delta_k > M_0 v_T^{-2}\}} |T^2(\delta_k v_T^3)^{-1} R_2(\delta_k, v_T)| \leq C$. To facilitate calculation, we further define a function $g: \mathbb{R}^3 \rightarrow \mathbb{R}$ such that

$$g(m_1, m_2, m_3) = m_1 d_{\beta,0}^2 \frac{\delta_k}{k} + m_2 d_{\beta,0}^2 \frac{\delta_k (2k + k_0)}{k^2} + m_3 d_{\beta,0}^2 \frac{\delta_k (2k + k_0)^2}{k^3}, \tag{2.93}$$

where $(m_1, m_2, m_3)' \in \mathbb{R}^3$. Clearly, g is linear. With the constructions of $\beta_{k,1}, \beta_{k,2}$ and g at hand, we are now ready to derive the expressions of $C_{k,i}$, $i = 2, 3, 4$. By somewhat

cumbersome calculation and the linearity of g , $C_{k,2}$ can be written as

$$\begin{aligned}
C_{k,2} &= \frac{1}{3}(\delta_k^3 v_T^6)^{-1} k^3 \beta_{k,2}^2 + (\delta_k^3 v_T^6)^{-1} k^2 \beta_{k,1} \beta_{k,2} + (\delta_k^3 v_T^6)^{-1} k \beta_{k,1}^2 + O(T^{-1}) \\
&= g\left(\frac{3}{4}, -4, \frac{4}{3}\right) + g\left(-\frac{3}{2}, 7, -2\right) + g(1, -4, 1) + O(T^{-1} v_T^{-2}) + O(T^{-1}) \\
&= g\left(\frac{1}{4}, -1, \frac{1}{3}\right) + O(T^{-1} v_T^{-2}) + O(T^{-1}).
\end{aligned} \tag{2.94}$$

Moreover, by the definition of $\beta_{k,1}$ and $\beta_{k,2}$, using Lemma 2.6, we get

$$\begin{aligned}
(\delta_k^3 v_T^6)^{-1} \beta_{k,1} \sum_{t=k_0+1}^k d_{k,t} &= (\delta_k^3 v_T^6)^{-1} \beta_{k,1} \left[d_{\mu,0} \delta_k v_T + \frac{1}{2} d_{\beta,0} \delta_k^2 v_T^3 + O(\delta_k v_T^3) \right] \\
&= \frac{1}{4} d_{\beta,0}^2 \frac{\delta_k}{k} - \frac{1}{2} d_{\beta,0}^2 \frac{\delta_k(2k+k_0)}{k^2} + O(T^{-1} v_T^{-2}), \\
(\delta_k^3 v_T^6)^{-1} \beta_{k,2} \sum_{t=k_0+1}^k t d_{k,t} &= (\delta_k^3 v_T^6)^{-1} \beta_{k,2} \left[\frac{1}{2} d_{\mu,0} \delta_k (k+k_0) v_T + \frac{1}{6} d_{\beta,0} \delta_k^2 (2k+k_0) v_T^3 + O(T \delta_k v_T^3) \right] \\
&= -\frac{1}{2} d_{\beta,0}^2 \frac{\delta_k(2k+k_0)}{k^2} + \frac{1}{3} d_{\beta,0}^2 \frac{\delta_k(2k+k_0)^2}{k^3} + O(T^{-1} v_T^{-2}).
\end{aligned}$$

Adding them together and multiply a factor 2 gives $C_{k,3} + C_{k,4} = g\left(\frac{1}{2}, -2, \frac{2}{3}\right) + O(T^{-1} v_T^{-2})$. As a result, together with (2.94), we obtain

$$C_{k,2} - C_{k,3} - C_{k,4} = g\left(-\frac{1}{4}, 1, -\frac{1}{3}\right) + O(T^{-1} v_T^{-2}) + O(T^{-1}). \tag{2.95}$$

Now it is easy to compute a lower bound of (2.95). Let $\pi = (2k+k_0)/k \in [2+\lambda_0/(1-\delta_2), 3)$. Then

$$\begin{aligned}
g\left(-\frac{1}{4}, 1, -\frac{1}{3}\right) &= d_{\beta,0}^2 \frac{\delta_k}{k} \left(-\frac{1}{3} \pi^2 + \pi - \frac{1}{4}\right) + O(T^{-1} v_T^{-2}) \\
&> -\frac{1}{4} d_{\beta,0}^2 \frac{\delta_k}{k} + O(T^{-1} v_T^{-2}) > -\frac{1}{4} d_{\beta,0}^2 + O(T^{-1} v_T^{-2}),
\end{aligned} \tag{2.96}$$

where $-\frac{1}{3} \pi^2 + \pi - \frac{1}{4}$ is decreasing and takes infimum at $\pi = 3$, and $\delta_k/k = 1 - k_0/k < 1$. Combine (2.91), (2.92), (2.95) and (2.96), $C_k > \frac{1}{12} d_{\beta,0}^2 + (\delta_k v_T^2)^{-1} d_{\mu,0} d_{\beta,0} + (\delta_k^2 v_T^4)^{-1} d_{\mu,0}^2 + O(v_T^2) + O(T^{-1} v_T^{-2})$. Let $l_0 = \frac{1}{12} d_{\beta,0}^2 > 0$, then $C_k \geq l_0 + O(M_0^{-1}) + O(M_0^{-2})$ when T is sufficiently large. If $d_{\beta,0} = 0$, let $C_0 = d_{\mu,0}^2 > 0$. ■

Appendix 2.B Infeasible Generalized Least Squares

We derive Propositions 1-2 as well as Theorems 2.1-2.2 in this section. These results are based on the intermediate lemmas in the previous section.

Proof of Proposition 1 Assume $d_{\beta,0} \neq 0$. To show Proposition 1, it suffices to argue that $\mathbb{P}(\hat{k} \in V_T(M_0))$ can be arbitrarily small as $T \rightarrow \infty$. Because $\text{SSR}(k_0; \widehat{\gamma}_{k_0}, \Sigma^{-1}(q)) \geq \text{SSR}(\hat{k}; \widehat{\gamma}_{\hat{k}}, \Sigma^{-1}(q))$ by definition, using (2.48) and Lemma 2.10, for every M_0 , there exist $T^* = T^*(M_0) > 0$ and $C_0 > 0$ such that

$$\begin{aligned}
 \{\hat{k} \in V_T(M_0)\} &\subset \left\{ \sup_{k \in V_T(M_0)} \left[\frac{\text{SSR}(k_0; \widehat{\gamma}_{k_0}, \Sigma^{-1}(q)) - \text{SSR}(k; \widehat{\gamma}_k, \Sigma^{-1}(q))}{\delta_k^3 v_T^6} \right] \geq 0 \right\} \\
 &\subset \left\{ \sup_{k \in V_T(M_0)} \left[\frac{-(XX)_k - 2(XE)_k - (EE)_k}{\delta_k^3 v_T^6} \right] \geq 0 \right\} \\
 &\subset \left\{ \sup_{k \in V_T(M_0)} \left[-\frac{(XX)_k}{\delta_k^3 v_T^6} \right] + \sup_{k \in V_T(M_0)} \left| \frac{2(XE)_k}{\delta_k^3 v_T^6} \right| + \sup_{k \in V_T(M_0)} \left| \frac{(EE)_k}{\delta_k^3 v_T^6} \right| \geq 0 \right\} \\
 &\subset \left\{ \sup_{k \in V_T(M_0)} \left| \frac{2(XE)_k}{\delta_k^3 v_T^6} \right| + \sup_{k \in V_T(M_0)} \left| \frac{(EE)_k}{\delta_k^3 v_T^6} \right| \geq \inf_{k \in V_T(M_0)} \frac{(XX)_k}{\delta_k^3 v_T^6} \right\} \\
 &\subset \left\{ \sup_{k \in V_T(M_0)} \left| \frac{2(XE)_k}{\delta_k^3 v_T^6} \right| + \sup_{k \in V_T(M_0)} \left| \frac{(EE)_k}{\delta_k^3 v_T^6} \right| \geq C_0 + O(M_0^{-1}) + O(M_0^{-2}) \right\} \\
 &=: E_T.
 \end{aligned} \tag{2.97}$$

It suffices to show, for every $\varepsilon > 0$, there exists a sufficiently large $M_0 = M_0(\varepsilon) > 0$ such that $\mathbb{P}(E_T) < \varepsilon, \forall T > T^*$. By the definition of $(XE)_k$ given in Lemma 2.4, and by Lemmas 2.5, 2.7 – 2.9, we have

$$\begin{aligned}
 \sup_{k \in V_T(M_0)} \left| \frac{(XE)_k}{\delta_k^3 v_T^6} \right| &\leq \sup_{k \in V_T(M_0)} \left| \frac{d'_k \Sigma^{-1}(q) \varepsilon}{\delta_k^3 v_T^6} \right| + \sup_{k \in V_T(M_0)} \left| \frac{1}{\delta_k^3 v_T^6} \frac{\delta_k^2 v_T^3}{T^{1/2}} \left[\left(\frac{\delta_k^2 v_T^3}{T^{1/2}} \right)^{-1} d'_k \Sigma^{-1}(q) X_k D_T^{-1/2} \right] \right. \\
 &\quad \left. \times \left[D_T^{-1/2} X'_k \Sigma^{-1}(q) X_k D_T^{-1/2} \right]^{-1} \left[D_T^{-1/2} X'_k \Sigma^{-1}(q) \varepsilon \right] \right| \\
 &= M_0^{-1} O_p(1) + O_p \left(\frac{\sqrt{\ln T}}{\sqrt{T} v_T} \right) = M_0^{-1} O_p(1).
 \end{aligned} \tag{2.98}$$

It remains to consider $(EE)_k$. Define $P_k = \Sigma^{-1}(q) X_k [X'_k \Sigma^{-1}(q) X_k]^{-1} X'_k \Sigma^{-1}(q)$, then $(EE)_k = \varepsilon' P_{k_0} \varepsilon - \varepsilon' P_k \varepsilon$. There are two key properties of P_k . First, it can be

written as

$$\mathbf{P}_k = \boldsymbol{\Sigma}^{-1/2}(q) \left\{ \boldsymbol{\Sigma}^{-1/2}(q) \mathbf{X}_k \left[\mathbf{X}_k' \boldsymbol{\Sigma}^{-1}(q) \mathbf{X}_k \right]^{-1} \mathbf{X}_k' \boldsymbol{\Sigma}^{-1/2}(q) \right\} \boldsymbol{\Sigma}^{-1/2}(q),$$

where the term within the curly brackets is a projection matrix. As a result, \mathbf{P}_k is positive semidefinite and has four nonzero eigenvalues. Second, suppose $\lambda_k = \lambda_{k,T} > 0$ is any nonzero eigenvalue of \mathbf{P}_k . Then $\lambda_u^{-1} \leq \lambda_k \leq \lambda_l^{-1}$, where λ_l and λ_u are positive constants that do not depend on k and T . To see this, we note that there exist $\lambda_l, \lambda_u > 0$ such that $\lambda_l \leq \lambda(\boldsymbol{\Sigma}(q)) \leq \lambda_u$, where $\lambda(\mathbf{A})$ denotes for some eigenvalue of a square matrix \mathbf{A} . Using equality $\mathbf{P}_k \boldsymbol{\Sigma}(q) \mathbf{P}_k = \mathbf{P}_k$, we deduce $\lambda_l \lambda_k^2 \leq \lambda_k \leq \lambda_u \lambda_k^2$ and then the result follows. By the spectral decomposition of \mathbf{P}_k , these two properties imply that $\boldsymbol{\varepsilon}' \mathbf{P}_k \boldsymbol{\varepsilon} = \boldsymbol{\varepsilon}' \left(\sum_{i=1}^4 \lambda_{i,k} \mathbf{q}_{i,k} \mathbf{q}_{i,k}' \right) \boldsymbol{\varepsilon} \leq \lambda_l^{-1} \sum_{i=1}^4 (\mathbf{q}_{i,k}' \boldsymbol{\varepsilon})^2$, $k \in \Gamma_T$, where $\mathbf{q}_{i,k}' \mathbf{q}_{j,k} = \mathbf{1}_{\{i=j\}}$. Obviously, $\mathbf{q}_{i,k}' \boldsymbol{\varepsilon}$ is a weighted sum of linear process $\{\epsilon_t\}$. Assumption 2.3 allows us to apply Lemma 6 (iv) of Dalla et al. (2014), and results in $\mathbb{E}(\mathbf{q}_{i,k}' \boldsymbol{\varepsilon})^2 = O(\|\mathbf{q}_{i,k}\|^2) = O(1)$. By the Chebychev inequality, $\sup_{k \in V_T(M_0)} \left| \frac{(EE)_k}{\delta_k^3 v_T^6} \right| < M_0^{-3} \sup_{k \in V_T(M_0)} |(EE)_k| = M_0^{-3} O_p(1)$. Using Equation (2.98), for every $\varepsilon > 0$, there exists a sufficiently large $M_0 > 0$ such that $\mathbb{P}(E_T) \leq \mathbb{P}(O_p(1) \geq M_0 C_0 + O(1)) < \varepsilon$ whenever $T > T^*$. If $d_{\beta,0} = 0$, we have $\sup_{k \in V_T(M_0)} \left| \frac{(XE)_k}{\delta_k v_T^2} \right| = O_p(1)$ and $\sup_{k \in V_T(M_0)} \left| \frac{(EE)_k}{\delta_k v_T^2} \right| = O_p(1)$. With Lemma 2.10 and some slight modifications, the result still holds. ■

Recall that the premultiplication by a lag polynomial should be understood as an operation on an element of a time series.

Proof of Theorem 2.1 Proposition 1 implies that \hat{k} will not lie in $V_T(M_0)$ with high probability. To study the limiting distribution of \hat{k} , we consider the behavior of $\text{SSR}(k; \widehat{\gamma}_k, \boldsymbol{\Sigma}^{-1}(q))$ on $K_T(M_0)$. By the definition of $(XX)_k - (EE)_k$ given in Lemma 2.4, together with Lemmas 2.5 and 2.7 – 2.9, we have

$$(XX)_k = \mathbf{d}_k' \boldsymbol{\Sigma}^{-1}(q) \mathbf{d}_k + o(1), \quad (XE)_k = \mathbf{d}_k' \boldsymbol{\Sigma}^{-1}(q) \boldsymbol{\varepsilon} + o_p(1) \quad \text{and} \quad (EE)_k = o_p(1) \quad (2.99)$$

where $o(1)$ and $o_p(1)$ are uniform on $K_T(M_0)$. Plug (2.99) in (2.48),

$$\text{SSR}(k_0; \widehat{\gamma}_{k_0}, \boldsymbol{\Sigma}^{-1}(q)) - \text{SSR}(k; \widehat{\gamma}_k, \boldsymbol{\Sigma}^{-1}(q)) = -\mathbf{d}_k' \boldsymbol{\Sigma}^{-1}(q) \mathbf{d}_k - 2\mathbf{d}_k' \boldsymbol{\Sigma}^{-1}(q) \boldsymbol{\varepsilon} + o_p(1), \quad (2.100)$$

where $o_p(1)$ is uniform on $K_T(M_0)$. For convenience, we let $\text{SSR}(k_0) - \text{SSR}(k)$ denote for the LHS of (2.100) from here on. Following an analogous reasoning as in the

proof of Proposition 3 in Bai (1997), for any given $M_0 > 0$, we assume $k = k_0 + \lfloor sv_T^{-2} \rfloor$, where $s \in [-M_0, M_0]$, and derive the corresponding limiting distribution of (2.100).

Consider $s \geq 0$ first. For the term $\mathbf{d}'_k \boldsymbol{\Sigma}^{-1}(q) \mathbf{d}_k$, by (2.62) and (2.63),

$$\begin{aligned} \mathbf{d}'_k \boldsymbol{\Sigma}^{-1}(q) \mathbf{d}_k &= \sigma^{-2}(q) A_q(1)^2 \sum_{t=k_0+1}^T \left(d_{k,t}^2 + 2d_{k,t} \tilde{d}_{k,t} + \tilde{d}_{k,t}^2 \right) \\ &= \sigma^{-2}(q) A_q(1)^2 \sum_{t=k_0+1}^T d_{k,t}^2 + O(v_T^2) + O(v_T^2) \\ &\xrightarrow{T \rightarrow \infty} \sigma_\epsilon^{-2} \left(\frac{1}{3} d_{\beta,0}^2 s^3 + d_{\mu,0} d_{\beta,0} s^2 + d_{\mu,0}^2 s \right), \end{aligned} \quad (2.101)$$

where $\left| \sum_{t=k_0+1}^T d_{k,t} \tilde{d}_{k,t} \right| \leq |\mu_{b,0}| \sum_{t=k_0+1}^k |\tilde{d}_{k,t}| + \delta_k |\beta_{b,0}| \sum_{t=k_0+1}^k \left| \frac{t-k_0}{\delta_k} \right| |\tilde{d}_{k,t}| = O(v_T^2)$ in the same fashion as in the proof of Lemma 2.7, and $\sum_{t=k_0+1}^T \tilde{d}_{k,t}^2 \leq \left(\sum_{t=k_0+1}^T |\tilde{d}_{k,t}| \right)^2 = O(v_T^2)$, and the last step is implied by Lemma 2.6. We move on to the term $\mathbf{d}'_k \boldsymbol{\Sigma}^{-1}(q) \boldsymbol{\epsilon}$. By straightforward calculation,

$$A_q(L) d_{k,t} = \left\{ \mu_{b,0} [A_q(L) \cdot 1] + \beta_{b,0} [A_q(L)(t - k_0)] - \tilde{A}_q(1) \beta_{b,0} \right\} \mathbf{1}_{\{k_0+1 \leq t \leq k\}} + A_q(1) \tilde{d}_{k,t}. \quad (2.102)$$

Then $\mathbf{d}'_k \boldsymbol{\Sigma}^{-1}(q) \boldsymbol{\epsilon}$ can be written as

$$\mathbf{d}'_k \boldsymbol{\Sigma}^{-1}(q) \boldsymbol{\epsilon} = \sigma^{-2}(q) \sum_{t=k_0+1}^T [A_q(L) d_{k,t}] [A_q(L) \epsilon_t] =: I_a + I_b - I_c + I_d, \quad (2.103)$$

where, by plugging in (2.102), we have $I_a = d_{\mu,0} v_T \sigma^{-2}(q) \sum_{t=k_0+1}^{k_0 + \lfloor sv_T^{-2} \rfloor} [A_q(L) \cdot 1] [A_q(L) \epsilon_t]$, $I_b = d_{\beta,0} v_T^3 \sigma^{-2}(q) \sum_{t=k_0+1}^{k_0 + \lfloor sv_T^{-2} \rfloor} [A_q(L)(t - k_0)] [A_q(L) \epsilon_t]$, and moreover, $I_c = d_{\beta,0} \tilde{A}_q(1) \sigma^{-2}(q) v_T^3 \sum_{t=k_0+1}^{k_0 + \lfloor sv_T^{-2} \rfloor} [A_q(L) \epsilon_t]$, $I_d = \sum_{t=k_0+1}^T [A_q(1) \tilde{d}_{k,t}] [A_q(L) \epsilon_t]$.

By the FCLTs (iii) and (iv) in Lemma 2.3, we have $I_a \xrightarrow{\mathcal{D}} \sigma_\epsilon^{-1} d_{\mu,0} W_1(s)$ and $I_b \xrightarrow{\mathcal{D}} \sigma_\epsilon^{-1} d_{\beta,0} \int_0^s r dW_1(r)$, where $W_1(r)$, $r \geq 0$, is a standard Brownian motion. Moreover, again by Lemma 2.3 (iii), we know $I_c = O_p(v_T^2)$. Finally, $I_d = O_p(v_T)$ using Cheby-

shev's inequality, (2.73) and the following results

$$\begin{aligned}
\mathbb{E} \left| \mu_{b,0} \sum_{j=1}^q \tilde{a}_j(q) \eta_{k_0+j}^* \right| &\leq |d_{\mu,0}| v_T \sum_{j=1}^q |\tilde{a}_j(q)| \sup_{t \in \mathbb{R}} |\eta_t^*| \leq C v_T, \\
\mathbb{E} \left| (\mu_{b,0} + \beta_{b,0} \delta_k) \sum_{j=1}^q \tilde{a}_j(q) \eta_{k+j}^* \right| &\leq (|d_{\mu,0}| + M_0 |d_{\beta,0}|) v_T \sum_{j=1}^q |\tilde{a}_j(q)| \sup_{t \in \mathbb{R}} |\eta_t^*| \leq C v_T, \\
\mathbb{E} \left| \beta_{b,0} \sum_{j=1}^q \tilde{a}_j(q) \sum_{t=k_0+j}^{k+j-1} \eta_t^* \right| &\leq |d_{\beta,0}| v_T^3 \sum_{j=1}^q |\tilde{a}_j(q)| \left(\delta_k \sup_{t \in \mathbb{R}} |\eta_t^*| \right) \leq C v_T.
\end{aligned} \tag{2.104}$$

Combining the results from (2.103) to (2.104), we arrive at

$$d'_k \Sigma^{-1}(q) \varepsilon \xrightarrow{\mathcal{D}} \sigma_\epsilon^{-1} \left(d_{\mu,0} W_1(s) + d_{\beta,0} \int_0^s r dW_1(r) \right), \quad s \geq 0. \tag{2.105}$$

The case $s < 0$ can be obtained similarly. We eventually get

$$\begin{aligned}
d'_k \Sigma^{-1}(q) d_k &\longrightarrow -\sigma_\epsilon^{-2} \left(\frac{1}{3} d_{\beta,0}^2 s^3 + d_{\mu,0} d_{\beta,0} s^2 + d_{\mu,0}^2 s \right) \\
d'_k \Sigma^{-1}(q) \varepsilon &\xrightarrow{\mathcal{D}} \sigma_\epsilon^{-1} \left(d_{\mu,0} W_2(-s) + d_{\beta,0} \int_0^{-s} r dW_2(r) \right)
\end{aligned} \tag{2.106}$$

where $W_2(-r)$, $r < 0$, is a standard Brownian motion that is independent of $W_1(s)$, $s \geq 0$. Summarizing (2.100), (2.101), (2.105) and (2.106), we have

$$\begin{aligned}
&\text{SSR}(k_0) - \text{SSR}(k_0 + [sv_T^{-2}]) \xrightarrow{\mathcal{D}} G(s) \\
&= \begin{cases} \sigma_\epsilon^{-2} \left(\frac{1}{3} d_{\beta,0}^2 s^3 + d_{\mu,0} d_{\beta,0} s^2 + d_{\mu,0}^2 s \right) - 2\sigma_\epsilon^{-1} \left(d_{\mu,0} W_2(-s) + d_{\beta,0} \int_0^{-s} r dW_2(r) \right), & s < 0, \\ -\sigma_\epsilon^{-2} \left(\frac{1}{3} d_{\beta,0}^2 s^3 + d_{\mu,0} d_{\beta,0} s^2 + d_{\mu,0}^2 s \right) - 2\sigma_\epsilon^{-1} \left(d_{\mu,0} W_1(s) + d_{\beta,0} \int_0^s r dW_1(r) \right), & s \geq 0, \end{cases}
\end{aligned}$$

where $s \in [-M_0, M_0]$ for any $M_0 > 0$. Moreover, it is not hard to obtain

$$G(s) \stackrel{\mathcal{D}}{=} \begin{cases} \mathbb{T}(s; d_{\mu,0}, d_{\beta,0}, \sigma_\epsilon^2) + 2W_2(\mathbb{T}(-s; d_{\mu,0}, d_{\beta,0}, \sigma_\epsilon^2)), & s < 0, \\ -\mathbb{T}(s; d_{\mu,0}, d_{\beta,0}, \sigma_\epsilon^2) + 2W_1(\mathbb{T}(s; d_{\mu,0}, d_{\beta,0}, \sigma_\epsilon^2)), & s \geq 0, \end{cases} \tag{2.107}$$

where \mathbb{T} is defined in (2.17) (the function is continuous and also monotone since $\mathbb{T}'(x; d_{\mu,0}, d_{\beta,0}, \sigma_\epsilon^2) = \sigma_\epsilon^{-2} (d_{\beta,0} x + d_{\mu,0})^2 \geq 0$).

Let $\hat{s}_T = v_T^2(\hat{k} - k_0)$. To obtain the distributional limit properties of \hat{s}_T , we are

going to apply the continuous mapping theorem for argmax-functional (Argmax-CMT) given in Corollary 5.58 of van der Vaart (1998). Note that $G(s)$ is a Gaussian process with continuous sample paths. By Lemma 2.6 of Kim and Pollard (1990), sample paths of $G(s)$ have unique maxima with probability 1. Moreover, $\hat{s}_T = \arg \max_{s \in [-M_0, M_0]} \{SSR(k_0) - SSR(k_0 + [sv_T^{-2}])\} + o_p(1) = O_p(1)$ is uniformly tight by Proposition 1. Applying the Argmax-CMT and the fact that $\lim_{|s| \rightarrow \infty} G(s) = -\infty$ almost surely, we obtain

$$\hat{s}_T \xrightarrow{\mathcal{D}} \arg \max_{s \in [-M_0, M_0]} G(s) = \arg \max_{s \in \mathbb{R}} G(s), \quad (2.108)$$

where the last equality is due to the fact that $\lim_{|s| \rightarrow \infty} G(s) = -\infty$ almost surely (which is true because the drift term dominates the random term in (2.107) when $|s|$ is large). Take $W(v) = W_1(v)$ for $v \geq 0$ and $W(v) = W_2(-v)$ for $v < 0$ as given in Theorem 2.1. By reparametrization, we obtain

$$\mathbb{T}(\hat{k} - k_0; \mu_{b,0}, \beta_{b,0}, \sigma_\epsilon^2) = \mathbb{T}(\hat{s}; d_{\mu,0}, d_{\beta,0}, \sigma_\epsilon^2) \xrightarrow{\mathcal{D}} \arg \max_{v \in \mathbb{R}} \left\{ W(v) - \frac{|v|}{2} \right\}, \quad (2.109)$$

which completes the proof. ■

Proof of Proposition 2 In Proposition 2, we compare the asymptotics of the aforementioned LS break date estimators. Recall $\Delta_\rho = \mathbf{I}_T - \rho \mathbf{L}_T$ defined in (2.19). It is noteworthy to mention that the asymptotic distribution of: (1) OLS estimator, which relies on $SSR(k; \hat{\gamma}_{k,OLS}, \mathbf{I}_T)$, assuming fixed break sizes is derived in Perron and Zhu (2005); (2) QD estimator, which is based on $SSR(k; \hat{\gamma}_{k,QD}, \Delta'_\rho \Delta_\rho)$, assuming Pitman drifts and AR(1) error process is developed in Harvey and Leybourne (2014). Since we adopt different assumptions, their limiting distributions are not directly comparable. To obtain the asymptotics, we essentially repeat the proofs starting from Lemma 2.4 for both estimators.

For a fixed $|\rho| < 1$, the asymptotic distribution of QD can be easily obtained by modifying the previous proofs. More specifically, it can be fulfilled by replacing $\Sigma^{-1}(q)$, $A_q(L)$, $\sigma^{-2}(q)$ and $\tilde{A}_q(L)$, respectively, by $\Delta'_\rho \Delta_\rho$, $1 - \rho L$, 1 and ρ . All the derived asymptotic orders remain the same. We omit the steps and list some examples below for illustration:

- (i) In Lemma 2.4, the three key terms have the identical formulations apart from we have $\Delta'_\rho \Delta_\rho$ instead of $\Sigma^{-1}(q)$ such as

$$(XX)_k = \mathbf{d}'_k \Delta'_\rho \left\{ \mathbf{I}_T - \Delta_\rho \mathbf{X}_k \left(\mathbf{X}'_k \Delta'_\rho \Delta_\rho \mathbf{X}_k \right)^{-1} \mathbf{X}'_k \Delta'_\rho \right\} \Delta_\rho \mathbf{d}_k;$$

(ii) In Lemma 2.5, (2.52) should be replaced by $\left[D_T^{-1/2} \mathbf{X}'_k \Delta'_\rho \Delta_\rho \mathbf{X}_k D_T^{-1/2} \right]^{-1} = (1 - \rho)^{-2} \Omega_\lambda^{-1} + O(T^{-1})$. The term $O(T^{-1})$ is again uniform over Γ_T ;

(iii) In Lemma 2.7, (2.59) should be replaced by

$$\begin{aligned} & D_T^{-1/2} \mathbf{X}'_k \Delta'_\rho \Delta_\rho \mathbf{d}_k \\ &= (1 - \rho)^2 D_T^{-1/2} \begin{pmatrix} \text{sgn}(k - k_0) d_{\mu,0} \delta_k v_T + \frac{1}{2} d_{\beta,0} \delta_k^2 v_T^3 \\ \text{sgn}(k - k_0) \frac{1}{2} d_{\mu,0} \delta_k (k + k_0) v_T + \frac{1}{6} d_{\beta,0} \delta_k^2 (2k + k_0) v_T^3 \\ \mathbf{1}_{\{k < k_0\}} \left(-d_{\mu,0} \delta_k v_T + \frac{1}{2} d_{\beta,0} \delta_k^2 v_T^3 \right) \\ \mathbf{1}_{\{k < k_0\}} \left(-\frac{1}{2} d_{\mu,0} \delta_k^2 v_T + \frac{1}{6} d_{\beta,0} \delta_k^3 v_T^3 \right) \end{pmatrix} + \mathbf{R}(\delta_k, v_T). \end{aligned}$$

Using these results, we eventually obtain

$$\begin{aligned} & (1 - \rho)^{-2} \sigma_\epsilon^{-2} \left\{ \text{SSR}(k_0; \widehat{\gamma}_{k_0, QD}, \Delta'_\rho \Delta_\rho) - \text{SSR}(k; \widehat{\gamma}_{k, QD}, \Delta'_\rho \Delta_\rho) \right\} \\ &= -(1 - \rho)^{-2} \sigma_\epsilon^{-2} \mathbf{d}'_k \Delta'_\rho \Delta_\rho \mathbf{d}_k - (1 - \rho)^{-2} \sigma_\epsilon^{-2} \mathbf{d}'_k \Delta'_\rho \Delta_\rho \boldsymbol{\varepsilon} + o_p(1) \\ &= \text{SSR}(k_0; \widehat{\gamma}_{k_0}, \Sigma^{-1}(q)) - \text{SSR}(k; \widehat{\gamma}_k, \Sigma^{-1}(q)) + o_p(1). \quad (2.110) \end{aligned}$$

In words, QD only differs from GLS with a constant scale $(1 - \rho)^{-2} \sigma_\epsilon^{-2}$ asymptotically. Taking $\rho = 0$, we immediately obtain the results of OLS. Clearly, they are asymptotically equivalent. \blacksquare

Proof of Theorem 2.2 Recall \mathbf{d}_k defined in (2.49), by adding and subtracting,

$$\begin{aligned} & D_T^{1/2} (\widehat{\gamma}_k - \gamma_0) \\ &= \left[D_T^{-1/2} \mathbf{X}'_k \Sigma^{-1}(q) \mathbf{X}_k D_T^{-1/2} \right]^{-1} \left[D_T^{-1/2} \mathbf{X}'_k \Sigma^{-1}(q) (\mathbf{X}_{k_0} - \mathbf{X}_k) \gamma_0 + D_T^{-1/2} \mathbf{X}'_k \Sigma^{-1}(q) \boldsymbol{\varepsilon} \right] \\ &= \left[D_T^{-1/2} \mathbf{X}'_k \Sigma^{-1}(q) \mathbf{X}_k D_T^{-1/2} \right]^{-1} \\ &\quad \times \left[D_T^{-1/2} \mathbf{X}'_k \Sigma^{-1}(q) \mathbf{d}_k + (\widehat{k} - k_0) D_T^{-1/2} \mathbf{X}'_k \Sigma^{-1}(q) \mathbf{Z}_k \gamma_0 + D_T^{-1/2} \mathbf{X}'_k \Sigma^{-1}(q) \boldsymbol{\varepsilon} \right]. \quad (2.111) \end{aligned}$$

Lemmas 2.5 and 2.7 imply

$$\left[D_T^{-1/2} X'_k \Sigma^{-1}(q) X_k D_T^{-1/2} \right]^{-1} \xrightarrow{\mathcal{P}} \sigma_\epsilon^2 \Omega_{\lambda_0}^{-1} \quad \text{and} \quad D_T^{-1/2} X'_k \Sigma^{-1}(q) d_k = o_p(1), \quad (2.112)$$

respectively. Furthermore, by Lemmas 2.3 and 2.9,

$$\begin{aligned} D_T^{-1/2} X'_k \Sigma^{-1}(q) \varepsilon &= D_T^{-1/2} X'_{k_0} \Sigma^{-1}(q) \varepsilon + o_p(1) \\ &\xrightarrow{\mathcal{D}} \sigma_\epsilon^{-1} \begin{pmatrix} W(1) \\ \int_0^1 r dW(r) \\ W(1) - W(\lambda_0) \\ \int_{\lambda_0}^1 (r - \lambda_0) dW(r) \end{pmatrix} =: \sigma_\epsilon^{-1} \mathbf{W}. \end{aligned} \quad (2.113)$$

It remains to consider the term $(\hat{k} - k_0) D_T^{-1/2} X'_k \Sigma^{-1}(q) Z_k \gamma_0$. Note that $Z_k \gamma_0 = \beta_{b,0} X_{\hat{k},3}$, where $X_{\hat{k},3}$ is the third column of X_k with elements $\mathbf{1}_{\{t > k\}}$. Hence,

$$(\hat{k} - k_0) D_T^{-1/2} X'_k \Sigma^{-1}(q) Z_k \gamma_0 = T^{1/2} \beta_{b,0} (\hat{k} - k_0) \left[T^{-1/2} D_T^{-1/2} X'_k \Sigma^{-1}(q) X_{\hat{k},3} \right]. \quad (2.114)$$

In a similar fashion as (2.54) and (2.55) in the proof of Lemma 2.5, using the BN decomposition,

$$\begin{aligned} T^{-1/2} D_T^{-1/2} X'_k \Sigma^{-1}(q) X_{\hat{k},3} &= \sigma^{-2}(q) A_q(1)^2 \begin{pmatrix} 1 - \hat{\lambda} \\ \frac{1 - \hat{\lambda}^2}{2} \\ 1 - \hat{\lambda} \\ \frac{(1 - \hat{\lambda})^2}{2} \end{pmatrix} + O_p(T^{-1}) \\ &= \sigma_\epsilon^{-2} \begin{pmatrix} 1 - \lambda_0 \\ \frac{1 - \lambda_0}{2} \\ 1 - \lambda_0 \\ \frac{(1 - \lambda_0)^2}{2} \end{pmatrix} + O_p(T^{-1} v_T^{-2}), \end{aligned} \quad (2.115)$$

where $\hat{\lambda} = \hat{k}/T$ as defined below Proposition 1. Note that $T^{1/2} \beta_{b,0} (\hat{k} - k_0) = O_p(T^{1/2} v_T)$.

Combine the results (2.111) – (2.115),

$$\begin{aligned}
D_T^{1/2}(\widehat{\gamma}_{\hat{k}} - \gamma_0) &= \Omega_{\lambda_0}^{-1} \left\{ T^{1/2} \beta_{b,0}(\hat{k} - k_0) \begin{pmatrix} 1 - \lambda_0 \\ \frac{1 - \lambda_0^2}{2} \\ 1 - \lambda_0 \\ \frac{(1 - \lambda_0)^2}{2} \end{pmatrix} + \sigma_\epsilon \begin{pmatrix} W(1) \\ \int_0^1 r dW(r) \\ W(1) - W(\lambda_0) \\ \int_{\lambda_0}^1 (r - \lambda_0) dW(r) \end{pmatrix} \right\} + o_p(1) \\
&= T^{1/2} \beta_{b,0}(\hat{k} - k_0) \begin{pmatrix} 0 \\ 0 \\ 1 \\ 0 \end{pmatrix} + \sigma_\epsilon \Omega_{\lambda_0}^{-1} \begin{pmatrix} W(1) \\ \int_0^1 r dW(r) \\ W(1) - W(\lambda_0) \\ \int_{\lambda_0}^1 (r - \lambda_0) dW(r) \end{pmatrix} + o_p(1),
\end{aligned} \tag{2.116}$$

where $\sigma_\epsilon \Omega_{\lambda_0}^{-1} \mathbf{W} \stackrel{\mathcal{D}}{=} \mathcal{N}(\mathbf{0}, \sigma_\epsilon^2 \Omega_{\lambda_0}^{-1})$, and using the fact that the covariance matrix of \mathbf{W} (defined in (2.113)) is Ω_{λ_0} . Equation (2.116) has several implications. First, (2.21) is obtained by a straightforward rewriting. Second, if $d_{\beta,0} = 0$, we see that $D_T^{1/2}(\widehat{\gamma}_{\hat{k}} - \gamma_0)$ and $D_T^{1/2}(\widehat{\gamma}_{k_0} - \gamma_0)$ share the asymptotic distribution $\mathcal{N}(\mathbf{0}, \sigma_\epsilon^2 \Omega_{\lambda_0}^{-1})$. It further implies the marginal distribution (2.22). Third, if $d_{\beta,0} \neq 0$ and by (2.108), we have $v_T^{-1} \beta_{b,0}(\hat{k} - k_0) = d_{\beta,0} v_T^2(\hat{k} - k_0) \xrightarrow{\mathcal{D}} d_{\beta,0} \arg \max_{s \in \mathbb{R}} G(s)$. Since $v_T^{-1}(\hat{\mu}_{b,\hat{k}} - \mu_{b,0}) - v_T^{-1} \beta_{b,0}(\hat{k} - k_0) = T^{-1/2} v_T^{-1} O_p(1) = o_p(1)$, then (2.23) follows. ■

Appendix 2.C Consistency of Estimated BIAM

This section is dedicated to the proof of Theorem 2.3. Theorem 2 in Lin and Reuvers (2020b) provides the consistency result with conditions that are easy to verify. However, according to Lemma 2.11 below, their Assumption 2 is violated and the result cannot be immediately applied.

Lemma 2.11

Under Assumptions 2.1 to 2.4,

$$\|\widehat{\varepsilon}(\hat{k}_r) - \varepsilon\|^2 = O_p(1) \mathbf{1}_{\{d_{\beta,0}=0\}} + O_p(T v_T^2) \mathbf{1}_{\{d_{\beta,0} \neq 0\}}. \tag{2.117}$$

Proof Repeated addition and subtraction imply

$$\|\widehat{\varepsilon}(\hat{k}_r) - \varepsilon\|^2 = \|\mathbf{X}_{\hat{k}_r} \widehat{\gamma}_{\hat{k}_r} - \mathbf{X}_{k_0} \gamma_0\|^2 \leq C(I + II + III), \quad (2.118)$$

where $I = \left\|(\mathbf{X}_{\hat{k}_r} - \mathbf{X}_{k_0})(\widehat{\gamma}_{\hat{k}_r} - \gamma_0)\right\|^2$, $II = \left\|(\mathbf{X}_{\hat{k}_r} - \mathbf{X}_{k_0})\gamma_0\right\|^2$ and $III = \left\|\mathbf{X}_{k_0}(\widehat{\gamma}_{\hat{k}_r} - \gamma_0)\right\|^2$. To derive the asymptotic orders of these terms, it is important to know the property of \hat{k}_r . As discussed below Corollary 6 of Harvey and Leybourne (2014), \hat{k}_r is asymptotically independent of $\hat{\rho}_r$. It is reasoned as follows. By (2.110),

$$\begin{aligned} (\hat{k}_r, \hat{\rho}_r) &= \arg \min_{(k, \bar{\rho}) \in \Gamma_T \times \mathcal{R}_m} \left\{ \text{SSR}\{k; \widehat{\gamma}(k; \Delta'_{\bar{\rho}} \Delta_{\bar{\rho}}), \Delta'_{\bar{\rho}} \Delta_{\bar{\rho}}\} - \text{SSR}(k_0; \widehat{\gamma}_{k_0}, \Sigma^{-1}(q)) \right\} \\ &= \arg \min_{(k, \bar{\rho}) \in \Gamma_T \times \mathcal{R}_m} \left\{ \left[\text{SSR}\{k; \widehat{\gamma}(k; \Delta'_{\bar{\rho}} \Delta_{\bar{\rho}}), \Delta'_{\bar{\rho}} \Delta_{\bar{\rho}}\} - \text{SSR}\{k_0; \widehat{\gamma}(k_0; \Delta'_{\bar{\rho}} \Delta_{\bar{\rho}}), \Delta'_{\bar{\rho}} \Delta_{\bar{\rho}}\} \right] \right. \\ &\quad \left. + \left[\text{SSR}\{k_0; \widehat{\gamma}(k_0; \Delta'_{\bar{\rho}} \Delta_{\bar{\rho}}), \Delta'_{\bar{\rho}} \Delta_{\bar{\rho}}\} - \text{SSR}(k_0; \widehat{\gamma}_{k_0}, \Sigma^{-1}(q)) \right] \right\} \\ &= \arg \min_{(k, \bar{\rho}) \in \Gamma_T \times \mathcal{R}_m} \left\{ -(1 - \bar{\rho})^2 \sigma_\varepsilon^2 \left[\text{SSR}(k_0; \widehat{\gamma}_{k_0}, \Sigma^{-1}(q)) - \text{SSR}(k; \widehat{\gamma}_k, \Sigma^{-1}(q)) \right] \right. \\ &\quad \left. + \left[\text{SSR}\{k_0; \widehat{\gamma}(k_0; \Delta'_{\bar{\rho}} \Delta_{\bar{\rho}}), \Delta'_{\bar{\rho}} \Delta_{\bar{\rho}}\} - \text{SSR}(k_0; \widehat{\gamma}_{k_0}, \Sigma^{-1}(q)) \right] \right\} + o_p(1). \end{aligned}$$

The term $(1 - \bar{\rho})^2 \sigma_\varepsilon^2$ is positive and independent of k , while $\text{SSR}(k_0; \widehat{\gamma}_{k_0}, \Sigma^{-1}(q)) - \text{SSR}(k; \widehat{\gamma}_k, \Sigma^{-1}(q))$ is independent of $\bar{\rho}$. Moreover, the second part measures the errors caused by difference between $\sigma_\varepsilon^{-2} \Delta'_{\bar{\rho}} \Delta_{\bar{\rho}}$ and $\Sigma^{-1}(q)$, which is obviously independent of k . For any $|\bar{\rho}| < 1$, the optimal value of k maximizes the infeasible GLS sum of squared errors function $\text{SSR}(k_0; \widehat{\gamma}_{k_0}, \Sigma^{-1}(q)) - \text{SSR}(k; \widehat{\gamma}_k, \Sigma^{-1}(q))$ asymptotically, and is thus independent of $\hat{\rho}_r$ and equivalent with other LS estimators by Proposition 2. Such equivalence gives $v_T^2(\hat{k}_r - k_0) = O_p(1)$ and further implies the asymptotic equivalence between $\widehat{\gamma}_{\hat{k}_r}$ and $\widehat{\gamma}_{\hat{k}}$. From Theorem 2.2, we conclude that $\mathbf{D}_T^{1/2}(\widehat{\gamma}_{\hat{k}_r} - \gamma_0) = O_p(1)$ if $d_{\beta,0} = 0$, and similarly, $\mathbf{D}_{v_T, T}^{1/2}(\widehat{\gamma}_{\hat{k}_r} - \gamma_0) = O_p(1)$ if $d_{\beta,0} \neq 0$, where $\mathbf{D}_{v_T, T} = \text{diag}(T, T^3, v_T^{-2}, T^3)$ is given in Remark 2.3. Now we discuss the terms in (2.118) for different cases of $d_{\beta,0}$. If $d_{\beta,0} = 0$, we have

$$\begin{aligned} I &\leq \left\|(\mathbf{X}_{\hat{k}_r} - \mathbf{X}_{k_0}) \mathbf{D}_T^{-1/2}\right\|^2 \left\|\mathbf{D}_T^{1/2}(\widehat{\gamma}_{\hat{k}_r} - \gamma_0)\right\|^2 = o_p(1) O_p(1) = o_p(1), \\ II &= d_{\mu,0}^2 v_T^2 |\hat{k}_r - k_0| = O_p(1), \\ III &\leq \left\|\mathbf{X}_{k_0} \mathbf{D}_T^{-1/2}\right\|^2 \left\|\mathbf{D}_T^{1/2}(\widehat{\gamma}_{\hat{k}_r} - \gamma_0)\right\|^2 = \left\|\mathbf{D}_T^{-1/2} \mathbf{X}_{k_0}' \mathbf{X}_{k_0} \mathbf{D}_T^{-1/2}\right\| \left\|\mathbf{D}_T^{1/2}(\widehat{\gamma}_{\hat{k}_r} - \gamma_0)\right\|^2 = O_p(1). \end{aligned}$$

If $d_{\beta,0} \neq 0$,

$$\begin{aligned} I &\leq \left\| (X_{\hat{k}_r} - X_{k_0}) D_{v_T, T}^{-1/2} \right\|^2 \left\| D_{v_T, T}^{1/2} (\widehat{\gamma}_{\hat{k}_r} - \gamma_0) \right\|^2 = O_p(1) O_p(1) = O_p(1), \\ II &= O_p \left(v_T^2 |\hat{k}_r - k_0| + v_T^4 |\hat{k}_r - k_0|^2 + v_T^6 |\hat{k}_r - k_0|^3 + T v_T^6 |\hat{k}_r - k_0|^2 \right) = O_p \left(T v_T^2 \right), \\ III &\leq \left\| X_{k_0} D_T^{-1/2} D_{v_T, T}^{-1/2} D_T^{1/2} \right\|^2 \left\| D_{v_T, T}^{1/2} (\widehat{\gamma}_{\hat{k}_r} - \gamma_0) \right\|^2 = O_p \left(T v_T^2 \right), \end{aligned}$$

where $\left\| (X_{\hat{k}_r} - X_{k_0}) D_{v_T, T}^{-1/2} \right\|^2 \leq v_T^2 |\hat{k}_r - k_0| + T^{-3} \left(T |\hat{k}_r - k_0|^2 + \sum_{t=1}^l \hat{k}_r^{-k_0} t^2 \right) = O_p(1) + o_p(1) = O_p(1)$ by using $v_T^2 (\hat{k}_r - k_0) = O_p(1)$ and the norm property $\|\cdot\| \leq \|\cdot\|_{\mathcal{F}}$. Combining these results in Equation (2.118), we obtain the lemma. \blacksquare

Proof of Theorem 2.3 Note that $\|\widehat{\Sigma}^{-1}(q, \hat{k}_r) - \Sigma^{-1}\| \leq \|\Sigma^{-1}(q) - \Sigma^{-1}\| + \|\widehat{\Sigma}^{-1}(q, \hat{k}_r) - \Sigma^{-1}(q)\|$. Since the conditions in Lemma 2 of Cheng et al. (2015) are implied by Assumption 2.3, we have $\|\Sigma^{-1}(q) - \Sigma^{-1}\| \leq C \sqrt{\sum_{j=q+1}^{\infty} |a_j| \sum_{j=q+1}^{\infty} j |a_j|} \rightarrow 0$ as $T \rightarrow \infty$.⁸ By adding and subtracting, the second term can be bounded as

$$\begin{aligned} \left\| \widehat{\Sigma}^{-1}(q, \hat{k}_r) - \Sigma^{-1}(q) \right\| &\leq \left\| \widehat{F}(q, \hat{k}_r) - F(q) \right\| \left\| \widehat{S}^{-1}(q, \hat{k}_r) \right\| \left\| \widehat{F}(q, \hat{k}_r) \right\| \\ &+ \left\| F(q) \right\| \left\| \widehat{S}^{-1}(q, \hat{k}_r) - S^{-1}(q) \right\| \left\| \widehat{F}(q, \hat{k}_r) \right\| + \left\| F(q) \right\| \left\| S^{-1}(q) \right\| \left\| \widehat{F}(q, \hat{k}_r) - F(q) \right\|. \end{aligned}$$

We prove $\|\widehat{F}(q, \hat{k}_r) - F(q)\| = o_p(1)$ (see (2.126) below) and $\|\widehat{S}^{-1}(q, \hat{k}_r) - S^{-1}(q)\| = o_p(1)$ subsequently. The remaining terms are bounded in probability, implying Theorem 2.3. As Equation (A.49) in Cheng et al. (2015), we could write

$$\left\| \widehat{F}(q, \hat{k}_r) - F(q) \right\| \leq C \sqrt{q} \max_{1 \leq \ell \leq q} \left\| \widehat{a}(\ell, \hat{k}_r) - a(\ell) \right\| \quad (2.119)$$

where $\widehat{a}(\ell, \hat{k}_r)$ and $a(\ell)$ are defined in (2.27) and (2.5), respectively. Note that

$$\left[\widehat{a}(\ell, \hat{k}_r) - a(\ell) \right]' \widehat{\varepsilon}_\ell(\ell, \hat{k}_r) = \widehat{\eta}_{\ell+1}(\ell, \hat{k}_r) - \widetilde{\eta}_{\ell+1}(\ell, \hat{k}_r),$$

where $\widehat{\varepsilon}_\ell(\ell, \hat{k}_r) = [\widehat{\varepsilon}_\ell(\hat{k}_r), \dots, \widehat{\varepsilon}_{\ell-\ell+1}(\hat{k}_r)]'$ is a ℓ -dimensional vector containing residu-

⁸The condition (2.12) in Lemma 2 of Cheng et al. (2015) requires $\sup_{-\infty < t < \infty} \mathbb{E}(\eta_t^S | \mathcal{F}_{t-1}) \leq C_S$, where $S > 4$. However, this lemma is later used to prove their Theorem 3 which replaces the original condition by $\mathbb{E}(\eta_t^{2r} | \mathcal{F}_{t-1}) \leq C_{2r}$ with $r \geq 2$. In other words, taking $r = 2$ (Assumption 2.3) is sufficient. One could also refer to Lemma S4 by Lin and Reuvers (2020b).

als in a reversed-order and

$$\widehat{\eta}_{l+1}(\ell, \hat{k}_r) = \widehat{\varepsilon}_{l+1}(\hat{k}_r) - \mathbf{a}(\ell)' \widehat{\varepsilon}_l(\ell, \hat{k}_r), \quad \widetilde{\eta}_{l+1}(\ell, \hat{k}_r) = \widehat{\varepsilon}_{l+1}(\hat{k}_r) - \widehat{\mathbf{a}}(\ell, \hat{k}_r)' \widehat{\varepsilon}_l(\ell, \hat{k}_r). \quad (2.120)$$

By the first-order condition (2.27), we obtain

$$\begin{aligned} & \left(\widehat{\mathbf{a}}(\ell, \hat{k}_r) - \mathbf{a}(\ell) \right)' \left(\frac{1}{T-\ell} \sum_{i=\ell}^{T-1} \widehat{\varepsilon}_i(\ell, \hat{k}_r) \widehat{\varepsilon}_i(\ell, \hat{k}_r)' \right) \\ &= \frac{1}{T-\ell} \sum_{i=\ell}^{T-1} \left(\widehat{\eta}_{l+1}(\ell, \hat{k}_r) - \widetilde{\eta}_{l+1}(\ell, \hat{k}_r) \right) \widehat{\varepsilon}_i(\ell, \hat{k}_r)' = \frac{1}{T-\ell} \sum_{i=\ell}^{T-1} \widehat{\eta}_{l+1}(\ell, \hat{k}_r) \widehat{\varepsilon}_i(\ell, \hat{k}_r)'. \end{aligned} \quad (2.121)$$

We now argue that the $\ell \times \ell$ matrix $\frac{1}{T-\ell} \sum_{i=\ell}^{T-1} \widehat{\varepsilon}_i(\ell, \hat{k}_r) \widehat{\varepsilon}_i(\ell, \hat{k}_r)'$ is asymptotically invertible with probability 1. It suffices to prove $\frac{1}{T-q} \sum_{i=q}^{T-1} \widehat{\varepsilon}_i(q, \hat{k}_r) \widehat{\varepsilon}_i(q, \hat{k}_r)'$ is asymptotically invertible. Define $\varepsilon_\ell(\ell) = (\varepsilon_\ell, \dots, \varepsilon_{\ell+1})'$. By triangular inequality

$$\left\| \frac{1}{T-q} \sum_{i=q}^{T-1} \widehat{\varepsilon}_i(q, \hat{k}_r) \widehat{\varepsilon}_i(q, \hat{k}_r)' - \mathbb{E}(\varepsilon_i(q) \varepsilon_i(q)') \right\| \leq I_a + I_b, \quad (2.122)$$

where

$$\begin{aligned} I_a &= \left\| \frac{1}{T-q} \sum_{i=q}^{T-1} \widehat{\varepsilon}_i(q, \hat{k}_r) \widehat{\varepsilon}_i(q, \hat{k}_r)' - \frac{1}{T-q} \sum_{i=q}^{T-1} \varepsilon_i(q) \varepsilon_i(q)' \right\| \\ &\leq C \left\{ \frac{1}{T-q} \sum_{i=q}^{T-1} \|\widehat{\varepsilon}_i(q, \hat{k}_r) - \varepsilon_i(q)\|^2 + \sqrt{\frac{1}{T-q} \sum_{i=q}^{T-1} \|\widehat{\varepsilon}_i(q, \hat{k}_r) - \varepsilon_i(q)\|^2 \frac{1}{T-q} \sum_{i=q}^{T-1} \|\varepsilon_i(q)\|^2} \right\} \end{aligned}$$

and $I_b = \left\| \frac{1}{T-q} \sum_{i=q}^{T-1} \varepsilon_i(q) \varepsilon_i(q)' - \mathbb{E}(\varepsilon_i(q) \varepsilon_i(q)') \right\|$. For I_a , we have

$$\begin{aligned} & \frac{1}{T-q} \sum_{i=q}^{T-1} \|\widehat{\varepsilon}_i(q, \hat{k}_r) - \varepsilon_i(q)\|^2 \\ &= \frac{1}{T-q} \sum_{i=q}^{T-1} \sum_{s=i-q+1}^i |\widehat{\varepsilon}_s(\hat{k}_r) - \varepsilon_s|^2 \leq \frac{q}{T-q} \|\widehat{\varepsilon}(\hat{k}_r) - \varepsilon\|^2, \end{aligned} \quad (2.123)$$

and $\frac{1}{T-q} \sum_{i=q}^{T-1} \|\varepsilon_i(q)\|^2 \leq \frac{q}{T-q} \|\varepsilon\|^2 = O_p(q)$ using Markov's inequality. By Lemma 2.11 and Assumption 2.5, $I_a = O_p\left(\frac{q}{\sqrt{T}}\right) \mathbf{1}_{\{d_{\beta,0}=0\}} + O_p(q\nu_T) \mathbf{1}_{\{d_{\beta,0} \neq 0\}} = o_p(1)$. Moreover, a

direct application of the First Moment Bound Theorem of Findley and Wei (1993) leads to $I_b = O_p\left(\frac{q}{\sqrt{T}}\right) = o_p(1)$. The detailed proof is omitted here as it can be also found in Lemma 2 of Ing and Wei (2003), (A.45) of Cheng et al. (2015) and a multivariate generalization in Lemma S2 of Lin and Reuvers (2020b). Plugging the results of I_a and I_b in (3.43), we see $\frac{1}{T-q} \sum_{t=q}^{T-1} \widehat{\varepsilon}_t(q, \hat{k}_r) \widehat{\varepsilon}_t(q, \hat{k}_r)' = \mathbb{E}(\varepsilon_t(q) \varepsilon_t(q)') + o_p(1)$, where $\mathbb{E}(\varepsilon_t(q) \varepsilon_t(q)')$ is a leading principal submatrix of Σ and is thus invertible. Moreover, we first make a claim for RHS in (2.121),

$$\max_{1 \leq \ell \leq q} \left\| \frac{1}{T-\ell} \sum_{t=\ell}^{T-1} \widehat{\eta}_{t+1}(\ell, \hat{k}_r) \widehat{\varepsilon}_t(\ell, \hat{k}_r)' \right\| = O_p\left(\frac{q}{\sqrt{T}}\right) \mathbf{1}_{\{d_{\beta,0}=0\}} + O_p(qv_T) \mathbf{1}_{\{d_{\beta,0} \neq 0\}}. \quad (2.124)$$

The proof is given below (2.126). Consequently, by the invertibility of $\frac{1}{T-\ell} \sum_{t=\ell}^{T-1} \widehat{\varepsilon}_t(\ell, \hat{k}_r) \widehat{\varepsilon}_t(\ell, \hat{k}_r)'$, (2.121) and (2.124),

$$\begin{aligned} & \max_{1 \leq \ell \leq q} \left\| \widehat{a}(\ell, \hat{k}_r) - a(\ell) \right\| \\ & \leq \left\| \left[\frac{1}{T-q} \sum_{t=q}^{T-1} \widehat{\varepsilon}_t(q, \hat{k}_r) \widehat{\varepsilon}_t(q, \hat{k}_r)' \right]^{-1} \right\| \max_{1 \leq \ell \leq q} \left\| \frac{1}{T-\ell} \sum_{t=\ell}^{T-1} \widehat{\eta}_{t+1}(\ell, \hat{k}_r) \widehat{\varepsilon}_t(\ell, \hat{k}_r)' \right\| \\ & = O_p\left(\frac{q}{\sqrt{T}}\right) \mathbf{1}_{\{d_{\beta,0}=0\}} + O_p(qv_T) \mathbf{1}_{\{d_{\beta,0} \neq 0\}}, \end{aligned} \quad (2.125)$$

and eventually by (2.119), (2.125) and Assumption 2.5,

$$\left\| \widehat{F}(q, \hat{k}_r) - F(q) \right\| = O_p\left(\sqrt{\frac{q^3}{T}}\right) \mathbf{1}_{\{d_{\beta,0}=0\}} + O_p\left(\sqrt{q^3 v_T^2}\right) \mathbf{1}_{\{d_{\beta,0} \neq 0\}} = o_p(1). \quad (2.126)$$

We now return to show the claim (2.124), note that $\max_{1 \leq \ell \leq q} \left\| \frac{1}{T-\ell} \sum_{t=\ell}^{T-1} \widehat{\eta}_{t+1}(\ell, \hat{k}_r) \widehat{\varepsilon}_t(\ell, \hat{k}_r)' \right\| \leq II_a + \dots + II_d$, where

$$\begin{aligned} II_a &= \max_{1 \leq \ell \leq q} \left\| \frac{1}{T-\ell} \sum_{t=\ell}^{T-1} \eta_{t+1}(\ell) \varepsilon_t(\ell)' \right\| \\ II_b &= \max_{1 \leq \ell \leq q} \left\| \frac{1}{T-\ell} \sum_{t=\ell}^{T-1} (\widehat{\eta}_{t+1}(\ell, \hat{k}_r) - \eta_{t+1}(\ell)) \varepsilon_t(\ell)' \right\| \\ II_c &= \max_{1 \leq \ell \leq q} \left\| \frac{1}{T-\ell} \sum_{t=\ell}^{T-1} \eta_{t+1}(\ell) (\widehat{\varepsilon}_t(\ell, \hat{k}_r) - \varepsilon_t(\ell))' \right\| \\ II_d &= \max_{1 \leq \ell \leq q} \left\| \frac{1}{T-\ell} \sum_{t=\ell}^{T-1} (\widehat{\eta}_{t+1}(\ell, \hat{k}_r) - \eta_{t+1}(\ell)) (\widehat{\varepsilon}_t(\ell, \hat{k}_r) - \varepsilon_t(\ell))' \right\| \end{aligned}$$

with $\eta_{t+1}(\ell) = \epsilon_{t+1} - \mathbf{a}(\ell)' \epsilon_t(\ell)$ similarly as (2.120), and $\epsilon_t(\ell)$ is defined above (3.43). By (A.31) in Cheng et al. (2015), which is based on Lemmas 3 and 4 of Ing and Wei (2003), we have $\mathbb{E} \left\| \frac{1}{T-\ell} \sum_{t=\ell}^{T-1} \eta_{t+1}(\ell) \epsilon_t(\ell)' \right\|^2 \leq C \frac{\ell}{T}$. By Chebyshev's inequality, $\forall \epsilon > 0$, there exists $\alpha_\epsilon > 0$ such that

$$\mathbb{P} \left(II_a \geq \alpha_\epsilon \frac{q}{\sqrt{T}} \right) \leq \frac{1}{\alpha_\epsilon^2} \frac{T}{q^2} \sum_{\ell=1}^q \mathbb{E} \left\| \frac{1}{T-\ell} \sum_{t=\ell}^{T-1} \eta_{t+1}(\ell) \epsilon_t(\ell)' \right\|^2 \leq \epsilon,$$

namely $II_a = O_p \left(\frac{q}{\sqrt{T}} \right)$. Furthermore, if we write

$$\begin{aligned} \widehat{\eta}_{t+1}(\ell, \hat{k}_r) - \eta_{t+1}(\ell) &= (\widehat{\epsilon}_{t+1}(\hat{k}_r) - \epsilon_{t+1}) - \mathbf{a}(\ell)' (\widehat{\epsilon}_t(\ell, \hat{k}_r) - \epsilon_t(\ell)) \\ &= [1, -\mathbf{a}(\ell)'] (\widehat{\epsilon}_{t+1}(\ell + 1, \hat{k}_r) - \epsilon_{t+1}(\ell + 1)), \end{aligned} \quad (2.127)$$

then by Cauchy-Schwarz inequality and the Baxter's inequality ($\max_{1 \leq \ell \leq q} \|\mathbf{a}(\ell)\|^2 \leq C$), we obtain

$$(II_b)^2 \leq C \left(\max_{1 \leq \ell \leq q} \frac{1}{T-\ell} \sum_{t=\ell}^{T-1} \|\widehat{\epsilon}_{t+1}(\ell + 1, \hat{k}_r) - \epsilon_{t+1}(\ell + 1)\|^2 \right) \left(\max_{1 \leq \ell \leq q} \frac{1}{T-\ell} \sum_{t=\ell}^{T-1} \|\epsilon_t(\ell)\|^2 \right).$$

Similarly as (2.123), we have

$$\begin{aligned} \max_{1 \leq \ell \leq q} \frac{1}{T-\ell} \sum_{t=\ell}^{T-1} \|\widehat{\epsilon}_{t+1}(\ell + 1, \hat{k}_r) - \epsilon_{t+1}(\ell + 1)\|^2 \\ \leq \max_{1 \leq \ell \leq q} \frac{\ell + 1}{T-\ell} \|\widehat{\epsilon}(\hat{k}_r) - \epsilon\|^2 \leq \frac{q+1}{T-q} \|\widehat{\epsilon}(\hat{k}_r) - \epsilon\|^2, \end{aligned} \quad (2.128)$$

and $\max_{1 \leq \ell \leq q} \frac{1}{T-\ell} \sum_{t=\ell}^{T-1} \|\epsilon_t(\ell)\|^2 \leq \frac{q}{T-q} \|\epsilon\|^2 = O_p(q)$. By Lemma 2.11, we have $II_b =$

$O_p \left(\frac{q}{\sqrt{T}} \right) \mathbf{1}_{\{d_{\beta,0}=0\}} + O_p(qv_T) \mathbf{1}_{\{d_{\beta,0} \neq 0\}}$. Similarly,

$$\begin{aligned} II_c &\leq C \sqrt{\max_{1 \leq \ell \leq q} \frac{1}{T-\ell} \sum_{t=\ell}^{T-1} \|\epsilon_{t+1}(\ell + 1)\|^2} \sqrt{\max_{1 \leq \ell \leq q} \frac{1}{T-\ell} \sum_{t=\ell}^{T-1} \|\widehat{\epsilon}_t(\ell, \hat{k}_r) - \epsilon_t(\ell)\|^2} \\ &= O_p \left(\frac{q}{\sqrt{T}} \right) \mathbf{1}_{\{d_{\beta,0}=0\}} + O_p(qv_T) \mathbf{1}_{\{d_{\beta,0} \neq 0\}}. \end{aligned}$$

Finally, using (2.127) and (2.128), we obtain $II_d = O_p \left(\frac{q}{T} \right) \|\widehat{\epsilon}(\hat{k}_r) - \epsilon\|^2 = O_p \left(\frac{q}{T} \right) \mathbf{1}_{\{d_{\beta,0}=0\}} + O_p(qv_T^2) \mathbf{1}_{\{d_{\beta,0} \neq 0\}}$. Combining II_a to II_d leads to (2.124).

Finally, we consider $\|\widehat{\mathbf{S}}^{-1}(q, \hat{k}_r) - \mathbf{S}^{-1}(q)\|$. Note that

$$\|\widehat{\mathbf{S}}(q, \hat{k}_r) - \mathbf{S}(q)\| = \max \left\{ \left| \widehat{\sigma}^2(0, \hat{k}_r) - \sigma^2(0) \right|, \max_{1 \leq \ell \leq q} \left| \widehat{\sigma}^2(\ell, \hat{k}_r) - \sigma^2(\ell) \right| \right\}.$$

By definition, the first term can be written as

$$\begin{aligned} \left| \widehat{\sigma}^2(0, \hat{k}_r) - \sigma^2(0) \right| &= \left| \frac{1}{T} \sum_{t=1}^T \hat{\epsilon}_t^2(\hat{k}_r) - \mathbb{E} \epsilon_t^2 \right| \\ &\leq \frac{1}{T} \left\| \widehat{\epsilon}(\hat{k}_r) - \epsilon \right\|^2 + 2 \left| \frac{1}{T} \sum_{t=1}^T (\hat{\epsilon}_t(\hat{k}_r) - \epsilon_t) \epsilon_t \right| + \left| \frac{1}{T} \sum_{t=1}^T \epsilon_t^2 - \mathbb{E} \epsilon_t^2 \right| \\ &= O_p(T^{-1/2}) \mathbf{1}_{\{d_{\beta,0}=0\}} + O_p(v_T) \mathbf{1}_{\{d_{\beta,0} \neq 0\}} = o_p(1). \end{aligned} \quad (2.129)$$

Moreover, as (A.53) in Cheng et al. (2015), using (2.120) and (2.121), we can write

$$\begin{aligned} \max_{1 \leq \ell \leq q} \left| \widehat{\sigma}^2(\ell, \hat{k}_r) - \sigma^2(\ell) \right| &= \max_{1 \leq \ell \leq q} \left| \frac{1}{T-\ell} \sum_{t=\ell}^{T-1} \widehat{\eta}_{t+1}(\ell, \hat{k}_r)^2 - \sigma^2(\ell) \right| \\ &= \max_{1 \leq \ell \leq q} \left| \left(\frac{1}{T-\ell} \sum_{t=\ell}^{T-1} \widehat{\eta}_{t+1}(\ell, \hat{k}_r)^2 - \sigma^2(\ell) \right) - (\widehat{a}(\ell, \hat{k}_r) - a(\ell))' \left(\frac{1}{T-\ell} \sum_{t=\ell}^{T-1} \widehat{\eta}_{t+1}(\ell, \hat{k}_r) \widehat{\epsilon}_t(\ell, \hat{k}_r) \right) \right| \\ &\leq \max_{1 \leq \ell \leq q} \left| \frac{1}{T-\ell} \sum_{t=\ell}^{T-1} \widehat{\eta}_{t+1}(\ell, \hat{k}_r)^2 - \sigma^2(\ell) \right| + C \max_{1 \leq \ell \leq q} \left\| \frac{1}{T-\ell} \sum_{t=\ell}^{T-1} \widehat{\eta}_{t+1}(\ell, \hat{k}_r) \widehat{\epsilon}_t(\ell, \hat{k}_r) \right\|^2. \end{aligned} \quad (2.130)$$

For the first term, by definition, $\max_{1 \leq \ell \leq q} \left| \frac{1}{T-\ell} \sum_{t=\ell}^{T-1} \widehat{\eta}_{t+1}(\ell, \hat{k}_r)^2 - \sigma^2(\ell) \right| \leq III_a + III_b$, where

$$III_a = \max_{1 \leq \ell \leq q} \left| \frac{1}{T-\ell} \sum_{t=\ell}^{T-1} (\widehat{\eta}_{t+1}(\ell, \hat{k}_r)^2 - \eta_{t+1}(\ell)^2) \right|, \quad III_b = \max_{1 \leq \ell \leq q} \left| \frac{1}{T-\ell} \sum_{t=\ell}^{T-1} \eta_{t+1}(\ell)^2 - \mathbb{E} \eta_{t+1}(\ell)^2 \right|.$$

We can use similar arguments for III_b and III_c below Equation (2.127) and obtain

$$\begin{aligned} III_a &\leq C \max_{1 \leq \ell \leq q} \left\{ \frac{1}{T-\ell} \sum_{t=\ell}^{T-1} (\widehat{\eta}_{t+1}(\ell, \hat{k}_r) - \eta_{t+1}(\ell))^2 \right. \\ &\quad \left. + \sqrt{\frac{1}{T-\ell} \sum_{t=\ell}^{T-1} (\widehat{\eta}_{t+1}(\ell, \hat{k}_r) - \eta_{t+1}(\ell))^2 \frac{1}{T-\ell} \sum_{t=\ell}^{T-1} \eta_{t+1}(\ell)^2} \right\} \\ &= O_p\left(\frac{q}{\sqrt{T}}\right) \mathbf{1}_{\{d_{\beta,0}=0\}} + O_p(qv_T) \mathbf{1}_{\{d_{\beta,0} \neq 0\}}. \end{aligned}$$

Moreover, using Lemma 6 of Ing and Wei (2005), we obtain $\mathbb{E} \left| \frac{1}{T-\ell} \sum_{t=\ell}^{T-1} \eta_{t+1}(\ell)^2 - \mathbb{E} \eta_{t+1}(\ell)^2 \right|^2 \leq C(T-\ell)^{-1}$.⁹ It leads to $III_b = O_p \left(\sqrt{\frac{q}{T}} \right)$ by Chebyshev's inequality. In summary, we have

$$\max_{1 \leq \ell \leq q} \left| \frac{1}{T-\ell} \sum_{t=\ell}^{T-1} \widehat{\eta}_{t+1}(\ell, \hat{k}_r)^2 - \sigma^2(\ell) \right| = O_p \left(\frac{q}{\sqrt{T}} \right) \mathbf{1}_{\{d_{\beta,0}=0\}} + O_p(qv_T) \mathbf{1}_{\{d_{\beta,0} \neq 0\}}. \quad (2.131)$$

The second term in (2.130) is $O_p \left(\frac{q^2}{T} \right) \mathbf{1}_{\{d_{\beta,0}=0\}} + O_p(q^2 v_T^2) \mathbf{1}_{\{d_{\beta,0} \neq 0\}}$ according to (2.124). Together with (2.131), from (2.130), we have

$$\max_{1 \leq \ell \leq q} |\widehat{\sigma}^2(\ell, \hat{k}_r) - \sigma^2(\ell)| = O_p \left(\frac{q}{\sqrt{T}} \right) \mathbf{1}_{\{d_{\beta,0}=0\}} + O_p(qv_T) \mathbf{1}_{\{d_{\beta,0} \neq 0\}} = o_p(1). \quad (2.132)$$

From (2.129) and (2.132), we see $\|\widehat{S}(q, \hat{k}_r) - S(q)\| = o_p(1)$. The bound of the difference between the inverses follows from $\|\widehat{S}^{-1}(q, \hat{k}_r) - S^{-1}(q)\| \leq \|\widehat{S}(q, \hat{k}_r) - S(q)\| \|\widehat{S}(q, \hat{k}_r)^{-1}\| \|S(q)^{-1}\| = o_p(1)$. ■

Appendix 2.D Feasible Generalized Least Squares

Proof of Proposition 3 We shall argue why the asymptotic results of infeasible GLS continue to hold for feasible GLS in the order of Proposition 1, Theorem 2.1, Proposition 2 and finally Theorem 2.2. Since

$$\hat{k}_f = \arg \max_{k \in \Gamma_T} \left\{ \text{SSR}(k_0; \widehat{\gamma}_{k_0,f}, \widehat{\Sigma}^{-1}(q, \hat{k}_r)) - \text{SSR}(k; \widehat{\gamma}_{k,f}, \widehat{\Sigma}^{-1}(q, \hat{k}_r)) \right\}, \quad (2.133)$$

we first decompose the difference between SSRs inside brackets into three key terms as in Lemma 2.4. For convenience, if not explicitly mentioned, a notation with a hat above indicates the feasible counterpart of infeasible one in previous sections. For instance, $(\widehat{XX})_k$, $(\widehat{XE})_k$ and $(\widehat{EE})_k$ are similarly defined as in Lemma 2.4 by replacing $\Sigma^{-1}(q)$ with $\widehat{\Sigma}^{-1}(q, \hat{k}_r)$ in the constructions. We then write the difference term in (2.133) into $(\widehat{XX})_k$, $(\widehat{XE})_k$ and $(\widehat{EE})_k$ in the similar way of Lemma 2.4.

In order to mimic the proof of Proposition 1, we shall first obtain the inclusion result in (2.97) for FGLS. This result requires Lemma 2.10 which is argued next. We observe

⁹Lemma 6 in Ing and Wei (2005) is valid for m.d.s. as assumed in Assumption 2.3, see their Remark 5. A multivariate inequality can be found in Lemma S2 of Lin and Reuvers (2020b).

that

$$(\widehat{XX})_k = \mathbf{d}'_k \widehat{\mathbf{M}}_k \mathbf{d}_k = \mathbf{d}'_k \mathbf{M}_k \mathbf{d}_k + \mathbf{d}'_k (\widehat{\mathbf{M}}_k - \mathbf{M}_k) \mathbf{d}_k = (XX)_k + \mathbf{d}'_k (\widehat{\mathbf{M}}_k - \mathbf{M}_k) \mathbf{d}_k,$$

where $\mathbf{M}_k := \Sigma^{-1}(q) - \Sigma^{-1}(q) \mathbf{X}_k [\mathbf{X}'_k \Sigma^{-1}(q) \mathbf{X}_k]^{-1} \mathbf{X}'_k \Sigma^{-1}(q)$ and $\widehat{\mathbf{M}}_k$ is the feasible counterpart. We would like to show $\mathbf{d}'_k (\widehat{\mathbf{M}}_k - \mathbf{M}_k) \mathbf{d}_k \leq \|\widehat{\mathbf{M}}_k - \mathbf{M}_k\| \|\mathbf{d}_k\|^2 = o_p(1) \|\mathbf{d}_k\|^2$. If this is true, then we divide $(\widehat{XX})_k$ by $\delta_k^3 v_T^6$ in both sides and apply (2.57) in Lemma 2.6, obtaining similar result for $(\widehat{XX})_k$ as in Lemma 2.10. It suffices to show $\|\widehat{\mathbf{M}}_k - \mathbf{M}_k\| = o_p(1)$ uniformly over $k \in \Gamma_T$ for $\mathbf{d}'_k (\widehat{\mathbf{M}}_k - \mathbf{M}_k) \mathbf{d}_k$. Triangular inequality implies $\|\widehat{\mathbf{M}}_k - \mathbf{M}_k\| \leq I_a + I_b$ with $I_a = \|\widehat{\Sigma}^{-1}(q, \hat{k}_r) - \Sigma^{-1}(q)\| = o_p(1)$ by Theorem 2.3. Moreover,

$$\begin{aligned} I_b &= \left\| \widehat{\Sigma}^{-1}(q, \hat{k}_r) \mathbf{X}_k [\mathbf{X}'_k \widehat{\Sigma}^{-1}(q, \hat{k}_r) \mathbf{X}_k]^{-1} \mathbf{X}'_k \widehat{\Sigma}^{-1}(q, \hat{k}_r) \right. \\ &\quad \left. - \Sigma^{-1}(q) \mathbf{X}_k [\mathbf{X}'_k \Sigma^{-1}(q) \mathbf{X}_k]^{-1} \mathbf{X}'_k \Sigma^{-1}(q) \right\| \\ &= \left\| \widehat{\Sigma}^{-1}(q, \hat{k}_r) \mathbf{X}_k \mathbf{D}_T^{-1/2} [\mathbf{D}_T^{-1/2} \mathbf{X}'_k \widehat{\Sigma}^{-1}(q, \hat{k}_r) \mathbf{X}_k \mathbf{D}_T^{-1/2}]^{-1} \mathbf{D}_T^{-1/2} \mathbf{X}'_k \widehat{\Sigma}^{-1}(q, \hat{k}_r) \right. \\ &\quad \left. - \Sigma^{-1}(q) \mathbf{X}_k \mathbf{D}_T^{-1/2} [\mathbf{D}_T^{-1/2} \mathbf{X}'_k \Sigma^{-1}(q) \mathbf{X}_k \mathbf{D}_T^{-1/2}]^{-1} \mathbf{D}_T^{-1/2} \mathbf{X}'_k \Sigma^{-1}(q) \right\|. \end{aligned}$$

By Theorem 2.3 and the inequality $\|A'BA\| \leq \|B\| \|A'A\|$ for a symmetric matrix B , we obtain

$$\begin{aligned} &\|\mathbf{D}_T^{-1/2} \mathbf{X}'_k \widehat{\Sigma}^{-1}(q, \hat{k}_r) - \mathbf{D}_T^{-1/2} \mathbf{X}'_k \Sigma^{-1}(q)\| \\ &\leq \|\widehat{\Sigma}^{-1}(q, \hat{k}_r) - \Sigma^{-1}(q)\| \|\mathbf{D}_T^{-1/2} \mathbf{X}'_k \mathbf{X}_k \mathbf{D}_T^{-1/2}\|^{1/2} = o_p(1) \end{aligned}$$

and

$$\begin{aligned} &\left\| \mathbf{D}_T^{-1/2} \mathbf{X}'_k \widehat{\Sigma}^{-1}(q, \hat{k}_r) \mathbf{X}_k \mathbf{D}_T^{-1/2} - \mathbf{D}_T^{-1/2} \mathbf{X}'_k \Sigma^{-1}(q) \mathbf{X}_k \mathbf{D}_T^{-1/2} \right\| \\ &\leq \left\| \widehat{\Sigma}^{-1}(q, \hat{k}_r) - \Sigma^{-1}(q) \right\| \left\| \mathbf{D}_T^{-1/2} \mathbf{X}'_k \mathbf{X}_k \mathbf{D}_T^{-1/2} \right\| = o_p(1), \quad (2.134) \end{aligned}$$

where both $o_p(1)$ terms are uniform on Γ_T . As a result, we have $I_b = o_p(1)$ uniformly. Combine the results of I_a and I_b for $\|\widehat{\mathbf{M}}_k - \mathbf{M}_k\|$, we deduce that

$$\|\widehat{\mathbf{M}}_k - \mathbf{M}_k\| = o_p(1), \quad \text{uniformly over } k \in \Gamma_T. \quad (2.135)$$

Therefore, Lemma 2.10 is valid similarly for $(\widehat{XX})_k$. With this validity, we find (2.97) also holds for FGLS.

With (2.97), we now need to show similar asymptotic results in (2.98) for FGLS. We first mimic (2.98). Let $\widehat{\mathbf{P}}_k = \widehat{\Sigma}^{-1}(q, \hat{k}_r) - \widehat{\mathbf{M}}_k$ be the feasible projection matrix. We would like to show

$$\sup_{k \in V_T(M_0)} \left| \frac{(\widehat{XE})_k}{\delta_k^3 v_T^6} \right| \leq \sup_{k \in V_T(M_0)} \left| \frac{\mathbf{d}'_k \widehat{\Sigma}^{-1}(q, \hat{k}_r) \varepsilon}{\delta_k^3 v_T^6} \right| + \sup_{k \in V_T(M_0)} \left| \frac{\mathbf{d}'_k \widehat{\mathbf{P}}_k \varepsilon}{\delta_k^3 v_T^6} \right| = O_p(1). \quad (2.136)$$

The first part in (2.136) can be obtained by modifying the proof of Lemma 2.8. The modification requires further notation $\widehat{A}_q(L) = 1 - \sum_{j=1}^q \hat{a}_j(q, \hat{k}_r) L^j$, where we suppress the dependence on \hat{k}_r without confusion. Similarly, for the case $k > k_0$,

$$\begin{aligned} \mathbf{d}'_k \widehat{\Sigma}^{-1}(q, \hat{k}_r) \varepsilon &= \widehat{\sigma}^{-2}(q, \hat{k}_r) \sum_{t=k_0+1}^T [\widehat{A}_q(L) d_{k,t}] [\widehat{A}_q(L) \varepsilon_t] \\ &= \widehat{\sigma}^{-2}(q, \hat{k}_r) \sum_{t=k_0+1}^T [A_q(L) d_{k,t} + (\widehat{A}_q(L) - A_q(L)) d_{k,t}] [\eta_t^* + (\widehat{A}_q(L) - A_q(L)) \varepsilon_t]. \end{aligned} \quad (2.137)$$

Note that $\widehat{\sigma}^2(q, \hat{k}_r) = \sigma^2(q) + o_p(1)$ by (2.132), and by $\|\cdot\|_1 \leq \sqrt{q} \|\cdot\|$ and (2.125),

$$\left| \widehat{A}_q(1) - A_q(1) \right| \leq \sum_{j=1}^q |\hat{a}_j(q, \hat{k}_r) - a_j(q)| \leq \sqrt{q} \|\widehat{\mathbf{a}}(q, \hat{k}_r) - \mathbf{a}(q)\| = o_p(1). \quad (2.138)$$

Consequently, $\sup_{k \in V_T(M_0)} \left| \frac{\mathbf{d}'_k \widehat{\Sigma}^{-1}(q, \hat{k}_r) \varepsilon}{\delta_k^3 v_T^6} \right| < M_0^{-1} \widehat{\sigma}^{-2}(q, \hat{k}_r) (II_a + \dots + II_d) = O_p(1)$, where we have $II_a = \sup_{k \in V_T(M_0)} \left| (\delta_k^2 v_T^4)^{-1} \sum_{t=k_0+1}^T [A_q(L) d_{k,t}] \eta_t^* \right| = O_p(1)$ by Lemma 2.8, and

$$\begin{aligned} II_b &= \sup_{k \in V_T(M_0)} \left| (\delta_k^2 v_T^4)^{-1} \sum_{t=k_0+1}^T [A_q(L) d_{k,t}] [(\widehat{A}_q(L) - A_q(L)) \varepsilon_t] \right| \\ &\leq |A_q(1)| \sum_{j=1}^q |\hat{a}_j(q, \hat{k}_r) - a_j(q)| \sup_{k \in V_T(M_0)} \left\{ (\delta_k^2 v_T^4)^{-1} \left| \sum_{t=k_0+1}^T d_{k,t} \varepsilon_{t-j} \right| + (\delta_k^2 v_T^4)^{-1} \left| \sum_{t=k_0+1}^T \widetilde{d}_{k,t} \varepsilon_{t-j} \right| \right\} \\ &= o_p(1) \end{aligned}$$

following from (2.138) and arguments for (2.70), (2.71). Moreover, let $\eta_t^* = 0$ if $t > T$

for convenience, then

$$\begin{aligned}
II_c &= \sup_{k \in V_T(M_0)} \left| (\delta_k^2 v_T^4)^{-1} \sum_{t=k_0+1}^T [(\widehat{A}_q(L) - A_q(L))d_{k,t}] \eta_t^* \right| \\
&= \sup_{k \in V_T(M_0)} \left| (\delta_k^2 v_T^4)^{-1} \sum_{j=1}^q (\hat{a}_j(q, \hat{k}_r) - a_j(q)) \sum_{t=k_0+1}^T d_{k,t} \eta_{t+j}^* \right| \\
&\leq \sum_{j=1}^q |\hat{a}_j(q, \hat{k}_r) - a_j(q)| \sup_{k \in V_T(M_0)} \left\{ (\delta_k^2 v_T^4)^{-1} \left| \sum_{t=k_0+1}^T d_{k,t} \eta_{t+j}^* \right| \right\} = o_p(1).
\end{aligned}$$

Similarly, we obtain $II_d = \sup_{k \in V_T(M_0)} |(\delta_k^2 v_T^4)^{-1} \sum_{t=k_0+1}^T [(\widehat{A}_q(L) - A_q(L))d_{k,t}] [(\widehat{A}_q(L) - A_q(L))\epsilon_t]| = o_p(1)$. The proof is similar for the case $k < k_0$.

For the second part in (2.136), by the trace version of Cauchy-Schwarz inequality $\text{tr}(\mathbf{A}'\mathbf{B})^2 \leq \text{tr}(\mathbf{A}\mathbf{A}'\mathbf{B}\mathbf{B}')$ (page 325, Abadir and Magnus (2005)), we have $|\mathbf{d}'_k \widehat{\mathbf{P}}_k \boldsymbol{\varepsilon}|^2 = \text{tr}(\widehat{\mathbf{P}}_k \boldsymbol{\varepsilon} \mathbf{d}'_k)^2 \leq \text{tr}(\widehat{\mathbf{P}}_k^2 \boldsymbol{\varepsilon} \mathbf{d}'_k \mathbf{d}_k \boldsymbol{\varepsilon})$, namely $|\mathbf{d}'_k \widehat{\mathbf{P}}_k \boldsymbol{\varepsilon}| \leq (\mathbf{d}'_k \mathbf{d}_k)^{1/2} (\boldsymbol{\varepsilon}' \widehat{\mathbf{P}}_k^2 \boldsymbol{\varepsilon})^{1/2}$. By the consistency of BIAM and Weyl's inequality (e.g. pages 40 and 46 in Tao (2012)), the eigenvalues of $\widehat{\boldsymbol{\Sigma}}(q, \hat{k}_r)$ are bounded from infinity and below away from zero in probability. Similar arguments below (2.98) imply $\boldsymbol{\varepsilon}' \widehat{\mathbf{P}}_k^2 \boldsymbol{\varepsilon} = O_p(1)$ uniformly over $k \in \Gamma_T$. Using Lemma 2.6 for $(\mathbf{d}'_k \mathbf{d}_k)^{1/2}$, we have $\sup_{k \in V_T(M_0)} \left| \frac{\mathbf{d}'_k \widehat{\mathbf{P}}_k \boldsymbol{\varepsilon}}{\delta_k^3 v_T^6} \right| = O_p(1)$ as expected for obtaining (2.136).

Finally, again using the property of bounded eigenvalues of $\widehat{\boldsymbol{\Sigma}}(q, \hat{k}_r)$, we see $\sup_{k \in V_T(M_0)} \left| \frac{(\widehat{EE})_k}{\delta_k^2 v_T^2} \right| = O_p(1)$ holds for $\boldsymbol{\varepsilon}' \widehat{\mathbf{P}}_{k_0} \boldsymbol{\varepsilon} - \boldsymbol{\varepsilon}' \widehat{\mathbf{P}}_k \boldsymbol{\varepsilon} \stackrel{\Delta}{=} (\widehat{EE})_k$. Then Proposition 1 is immediately valid for \hat{k}_f by (2.136).

We continue to argue that \hat{k}_f has the limiting distribution in Theorem 2.1. Similarly, we shall consider the properties of three key components $(\widehat{XX})_k$, $(\widehat{XE})_k$ and $(\widehat{EE})_k$. Remember we have

$$(\widehat{XX})_k = (XX)_k + o_p(1) \|\mathbf{d}_k\|^2 = \mathbf{d}'_k \boldsymbol{\Sigma}^{-1}(q) \mathbf{d}_k + o_p(1) \quad (2.139)$$

uniformly over $k \in K_T(M_0)$ by (2.99). Next, we show $(\widehat{XE})_k = \mathbf{d}'_k \boldsymbol{\Sigma}^{-1}(q) \boldsymbol{\varepsilon} + o_p(1)$ uniform on $K_T(M_0)$. By (2.137), (2.138) and Lemma 2.1, straightforward modifications of II_a to II_d below (2.138) imply $\mathbf{d}'_k \widehat{\boldsymbol{\Sigma}}^{-1}(q, \hat{k}_r) \boldsymbol{\varepsilon} = \mathbf{d}'_k \boldsymbol{\Sigma}^{-1}(q) \boldsymbol{\varepsilon} + o_p(1)$. Moreover,

because

$$\begin{aligned} & \left\| D_T^{-1/2} \mathbf{X}'_k (\widehat{\Sigma}^{-1}(q, \hat{k}_r) - \Sigma^{-1}(q)) \mathbf{d}_k \right\| \\ & \leq \| D_T^{-1/2} \mathbf{X}'_k \mathbf{X}_k D_T^{-1/2} \|^{1/2} \left\| \widehat{\Sigma}^{-1}(q, \hat{k}_r) - \Sigma^{-1}(q) \right\| \|\mathbf{d}_k\| = o_p(1), \end{aligned} \quad (2.140)$$

we obtain $D_T^{-1/2} \mathbf{X}'_k \widehat{\Sigma}^{-1}(q, \hat{k}_r) \mathbf{d}_k = D_T^{-1/2} \mathbf{X}'_k \Sigma^{-1}(q) \mathbf{d}_k + o_p(1) = o_p(1)$ by Lemma 2.7. Furthermore, as (2.81) in Lemma 2.9, by the BN decomposition of $\widehat{A}_q(L)$ and (2.134)

$$\begin{aligned} & D_T^{-1/2} \mathbf{X}'_k \widehat{\Sigma}^{-1}(q, \hat{k}_r) \varepsilon \\ & = \widehat{\sigma}^{-2}(q, \hat{k}_r) \begin{pmatrix} \frac{1}{T^{1/2}} \sum_{t=1}^T [\widehat{A}_q(L) \cdot 1] \\ \frac{1}{T^{3/2}} \sum_{t=1}^T [\widehat{A}_q(L) \cdot t] \\ \frac{1}{T^{1/2}} \sum_{t=1}^T [\widehat{A}_q(L) \mathbf{1}_{\{t > k\}}] \\ \frac{1}{T^{3/2}} \sum_{t=1}^T [\widehat{A}_q(L)(t - k) \mathbf{1}_{\{t > k\}}] \end{pmatrix} [\eta_t^* + (\widehat{A}_q(L) - A_q(L)) \varepsilon_t] \\ & = D_T^{-1/2} \mathbf{X}'_k \Sigma^{-1}(q) \varepsilon + o_p(1) = O_p(1) \end{aligned} \quad (2.141)$$

for any $k \in K_T(M_0)$. Using the results above and (2.134), we eventually have

$$\begin{aligned} (\widehat{XE})_k &= \mathbf{d}'_k \widehat{\Sigma}^{-1}(q, \hat{k}_r) \varepsilon - \mathbf{d}'_k \widehat{\Sigma}^{-1}(q, \hat{k}_r) \mathbf{X}_k D_T^{-1/2} \\ & \quad \times \left[D_T^{-1/2} \mathbf{X}'_k \widehat{\Sigma}^{-1}(q, \hat{k}_r) \mathbf{X}_k D_T^{-1/2} \right]^{-1} D_T^{-1/2} \mathbf{X}'_k \widehat{\Sigma}^{-1}(q, \hat{k}_r) \varepsilon = \mathbf{d}'_k \Sigma^{-1}(q) \varepsilon + o_p(1). \end{aligned} \quad (2.142)$$

By (2.134) and

$$\begin{aligned} & D_T^{-1/2} \mathbf{X}'_{k_0} \widehat{\Sigma}^{-1}(q, \hat{k}_r) \varepsilon - D_T^{-1/2} \mathbf{X}'_k \widehat{\Sigma}^{-1}(q, \hat{k}_r) \varepsilon \\ & = \widehat{\sigma}^{-2}(q, \hat{k}_r) \widehat{A}_q(1) \begin{pmatrix} \mathbf{0}_{2 \times 1} \\ T^{-1/2} \sum_{t=k_0+1}^k [\widehat{A}_q(L) \varepsilon_t] \\ T^{-3/2} \sum_{t=k_0+1}^k (t - k_0) [\widehat{A}_q(L) \varepsilon_t] \end{pmatrix} + o_p(1) = o_p(1), \end{aligned}$$

we find $(\widehat{EE})_k = o_p(1)$. With (2.139) and (2.142), we finally deduce that

$$\begin{aligned} & \text{SSR}(k_0; \widehat{\gamma}_{k_0, f}, \widehat{\Sigma}^{-1}(q, \hat{k}_r)) - \text{SSR}(k; \widehat{\gamma}_{k, f}, \widehat{\Sigma}^{-1}(q, \hat{k}_r)) \\ & = -\mathbf{d}'_k \Sigma^{-1}(q) \mathbf{d}_k - 2\mathbf{d}'_k \Sigma^{-1}(q) \varepsilon + o_p(1) \end{aligned} \quad (2.143)$$

which is equivalent with (2.100). Then Theorem 2.1 follows. Proposition 2 (asymptotic equivalence) is a natural consequence of Theorem 2.1. Finally, by (2.111), (2.114), (2.134), (2.140) and (2.141), we obtain $D_T^{1/2}(\widehat{\gamma}_{\hat{k}_f, f} - \gamma_0) = D_T^{1/2}(\widehat{\gamma}_{\hat{k}} - \gamma_0) + o_p(1)$. Hence, Theorem 2.2 follows for $\widehat{\gamma}_{\hat{k}_f, f}$. ■

Proof of Theorem 2.4 Recall the asymptotic equivalence between the difference of infeasible SSR (2.100) and feasible SSR (2.143), the cubic polynomial transformation \mathbb{T} in (2.17) and $G(s)$ in (2.107). By the definition of log-LR test and the continuous mapping theorem,

$$\begin{aligned} \text{LR}_T(k_0) &= \text{SSR}(k_0; \widehat{\gamma}_{k_0, f}, \widehat{\Sigma}^{-1}(q, \hat{k}_r)) - \text{SSR}(\hat{k}_f; \widehat{\gamma}_{\hat{k}_f, f}, \widehat{\Sigma}^{-1}(q, \hat{k}_r)) \\ &= \max_{k \in \Gamma_T} \left\{ \text{SSR}(k_0; \widehat{\gamma}_{k_0, f}, \widehat{\Sigma}^{-1}(q, \hat{k}_r)) - \text{SSR}(k; \widehat{\gamma}_{k, f}, \widehat{\Sigma}^{-1}(q, \hat{k}_r)) \right\} \\ &= \max_{s \in [-M_0, M_0]} \left\{ \text{SSR}(k_0) - \text{SSR}(k_0 + [sv_T^{-2}]) \right\} + o_p(1) \\ &\xrightarrow{\mathcal{D}} \max_{s \in [-M_0, M_0]} G(s) = \max_{s \in \mathbb{R}} G(s) = \max_{v \in \mathbb{R}} \{2W(v) - |v|\}, \end{aligned}$$

where the last step follows from the strict monotonicity of \mathbb{T} . ■

Appendix 2.E A Hájek and Rényi Inequality

The following result is a simple extension of Hájek and Rényi (HR) type of inequality (see Proposition 1 of Bai (1994)) to allow for time trend.

Lemma 2.12 (HR Inequality)

Consider a linear process $\varepsilon_t = C(L)\eta_t$, where $C(L) = \sum_{j=0}^{\infty} c_j L^j$ with $c_0 = 1$ and $\sum_{j=0}^{\infty} j|c_j| < \infty$ and $C(1) \neq 0$, and $\{\eta_t\}$ is a martingale difference sequence with $\mathbb{E}(\eta_t | \mathcal{F}_{t-1}) = 0$ and $\mathbb{E}(\eta_t^2 | \mathcal{F}_{t-1}) = \sigma^2$ almost surely. Let $K_j > 0$, $j = 1, 2, \dots$, be a sequence of non-increasing sequence. For any $\alpha > 0$,

$$\mathbb{P} \left(\max_{n \leq j \leq m} K_j \left| \sum_{t=1}^j t \varepsilon_t \right| > \alpha \right) \leq C_0 \frac{\sigma^2}{\alpha^2} \left(K_n^2 \sum_{t=1}^n t^2 + \sum_{t=n+1}^m t^2 K_t^2 \right), \quad (2.144)$$

where $C_0 < \infty$ is a constant that only depends on c_j 's.

Proof By the Beveridge-Nelson decomposition (Lemma 2.1 in Phillips and Solo (1992)), we can express $C(L) = C(1) - (1 - L)\tilde{C}(L)$, where $\tilde{C}(L) = \sum_{j=0}^{\infty} \tilde{c}_j L^j$ with $\tilde{c}_j = \sum_{k=j+1}^{\infty} c_k$ and $\sum_{j=0}^{\infty} |\tilde{c}_j| < \infty$. Let $\tilde{\eta}_t = \tilde{C}(L)\eta_t$. Since $t\Delta\tilde{\eta}_t = \Delta(t\tilde{\eta}_t) - \tilde{\eta}_{t-1}$, then $\left| \sum_{t=1}^j t\varepsilon_t \right| \leq \left| C(1) \sum_{t=1}^j t\eta_t \right| + j|\tilde{\eta}_j| + \left| \sum_{t=1}^j \tilde{\eta}_{t-1} \right|$. By the HR inequality for martingale differences,

$$\mathbb{P} \left(\max_{n \leq j \leq m} K_j \left| C(1) \sum_{t=1}^j t\eta_t \right| > \frac{\alpha}{3} \right) \leq \frac{9\sigma^2 C(1)^2}{\alpha^2} \left(K_n^2 \sum_{t=1}^n t^2 + \sum_{t=n+1}^m t^2 K_t^2 \right).$$

By Chebyshev's inequality,

$$\mathbb{P} \left(\max_{n \leq j \leq m} K_j j |\tilde{\eta}_j| > \frac{\alpha}{3} \right) \leq \sum_{t=n}^m \mathbb{P} \left(K_t t |\tilde{\eta}_t| > \frac{\alpha}{3} \right) \leq \frac{9\sigma^2 (\sum_{j=0}^{\infty} |\tilde{c}_j|)^2}{\alpha^2} \sum_{t=n}^m t^2 K_t^2,$$

where $\mathbb{E}\tilde{\eta}_t^2 \leq \sigma^2 (\sum_{j=0}^{\infty} |\tilde{c}_j|)^2$ by Minkowski's inequality. Similarly, by Chebyshev's inequality and c_r -inequality, we again obtain

$$\mathbb{P} \left(\max_{n \leq j \leq m} K_j \left| \sum_{t=1}^j \tilde{\eta}_{t-1} \right| > \frac{\alpha}{3} \right) \leq \frac{9}{\alpha^2} \sum_{j=n}^m j K_j^2 \sum_{t=1}^j \mathbb{E}\tilde{\eta}_{t-1}^2 \leq \frac{9\sigma^2 (\sum_{j=0}^{\infty} |\tilde{c}_j|)^2}{\alpha^2} \sum_{t=n}^m t^2 K_t^2.$$

Take $C_0 = 27 \max \{C(1)^2, (\sum_{j=0}^{\infty} |\tilde{c}_j|)^2\}$, then the lemma follows. \blacksquare

Corollary 2.3

Let $K_j = j^{-2}$ and $m \rightarrow \infty$, we have

$$\mathbb{P} \left(\sup_{j \geq n} \frac{1}{j} \left| \sum_{t=1}^j \frac{t}{j} \varepsilon_t \right| > \alpha \right) \leq C_0 \frac{\sigma^2}{\alpha^2} \left(\frac{1}{n^4} \sum_{t=1}^n t^2 + \frac{1}{n} \right) \leq \frac{C}{\alpha^2} \frac{1}{n}.$$

Similarly, take $K_j = j^{-3/2}$, then

$$\mathbb{P} \left(\max_{n \leq j \leq m} \frac{1}{\sqrt{j}} \left| \sum_{t=1}^j \frac{t}{j} \varepsilon_t \right| > \alpha \right) \leq C_0 \frac{\sigma^2}{\alpha^2} \left(\frac{1}{3} + \sum_{t=n}^m \frac{1}{t} \right) + O(n^{-1}).$$

Appendix 2.F Additional Simulation Outputs

Table 2.3: The empirical Mean Squared Error (MSE) and concentration probability of LS break date estimators with small breaks $(\mu_{\delta,0}, \beta_{\delta,0}) = (0.5, 0.05)$. The MSEs are expressed relative to QD. A value that is *larger* than 1 signifies that QD performs better. The raw outputs of QD are listed. The concentration probability measures the probability of a break fraction estimator to lie in the interval $[\lambda_0 - c, \lambda_0 + c]$.

$(\mu_{\delta,0}, \beta_{\delta,0}) = (0.5, 0.05)$																

Table 2.4: The empirical Mean Squared Error (MSE) and concentration probability of LS break date estimators with large breaks $(\mu_{b,0}, \beta_{b,0}) = (1, 0, 1)$. The MSEs are expressed relative to QD. A value that is *larger* than 1 signifies that QD performs better. The raw outputs of QD are listed. The concentration probability measures the probability of a break fraction estimator to lie in the interval $[\lambda_0 - c, \lambda_0 + c]$.

$(\mu_{b,0}, \beta_{b,0}) = (1, 0, 1)$		$\theta_0 = 0.3$					$\theta_0 = 0.6$					$\theta_0 = 0.9$				
		FGLS _{prebreak}	FGLS	OLS	QD	infeasible GLS	FGLS _{prebreak}	FGLS	OLS	QD	infeasible GLS	FGLS _{prebreak}	FGLS	OLS	QD	infeasible GLS
<i>T = 150, MSE Relative to QD</i>																
P_0	0.5	1.008	0.930	1.590	1.87E+01	0.799	0.643	0.552	1.233	4.49E+01	0.344	0.290	0.273	0.991	8.34E+01	0.023
	0.7	0.937	0.831	1.561	5.47E+01	0.640	0.580	0.448	1.235	1.33E+02	0.253	0.310	0.249	0.940	3.01E+02	0.003
	0.9	0.843	0.902	1.620	4.15E+02	0.739	0.524	0.585	1.155	7.73E+02	0.205	0.282	0.439	0.906	1.23E+03	0.000
P_0	0.95	0.828	0.946	1.511	6.57E+02	0.751	0.474	0.661	1.100	1.10E+03	0.216	0.289	0.553	0.880	1.52E+03	0.000
<i>T = 150, Concentration Probability with $c = 0.01$</i>																
P_0	0.5	0.839	0.839	0.729	0.834	0.857	0.830	0.834	0.580	0.738	0.868	0.847	0.843	0.466	0.620	0.926
	0.7	0.786	0.788	0.508	0.764	0.817	0.818	0.816	0.372	0.663	0.879	0.849	0.824	0.269	0.519	0.956
	0.9	0.727	0.685	0.222	0.662	0.758	0.782	0.717	0.153	0.537	0.881	0.825	0.709	0.122	0.394	0.979
P_0	0.95	0.697	0.653	0.167	0.617	0.737	0.765	0.643	0.115	0.485	0.883	0.816	0.624	0.088	0.350	0.980
<i>T = 150, Concentration Probability with $c = 0.03$</i>																
P_0	0.5	0.898	0.903	0.844	0.896	0.915	0.880	0.890	0.727	0.809	0.921	0.920	0.918	0.634	0.716	0.995
	0.7	0.823	0.831	0.640	0.805	0.854	0.853	0.857	0.510	0.718	0.912	0.897	0.874	0.408	0.586	0.998
	0.9	0.751	0.713	0.315	0.691	0.780	0.800	0.739	0.245	0.571	0.902	0.850	0.733	0.202	0.432	1.000
P_0	0.95	0.718	0.658	0.249	0.643	0.757	0.780	0.663	0.192	0.511	0.898	0.835	0.646	0.159	0.386	1.000
<i>T = 300, MSE Relative to QD</i>																
P_0	0.5	0.976	0.889	1.576	1.77E+01	0.843	0.493	0.459	1.593	3.19E+01	0.403	0.125	0.142	1.311	5.78E+01	0.023
	0.7	0.904	0.828	2.175	3.65E+01	0.769	0.398	0.355	1.866	6.67E+01	0.289	0.098	0.100	1.502	1.13E+02	0.008
	0.9	0.885	0.714	2.026	1.85E+02	0.584	0.364	0.318	1.311	5.23E+02	0.166	0.072	0.121	0.974	1.14E+03	0.000
P_0	0.95	0.854	0.814	2.063	6.42E+02	0.652	0.410	0.511	1.379	1.53E+03	0.216	0.107	0.319	1.083	2.75E+03	0.000
<i>T = 300, Concentration Probability with $c = 0.01$</i>																
P_0	0.5	0.884	0.890	0.802	0.877	0.897	0.906	0.907	0.694	0.813	0.919	0.961	0.958	0.580	0.709	0.986
	0.7	0.837	0.845	0.598	0.817	0.857	0.893	0.895	0.470	0.728	0.914	0.960	0.957	0.373	0.613	0.991
	0.9	0.784	0.781	0.258	0.730	0.804	0.876	0.847	0.181	0.615	0.906	0.961	0.900	0.134	0.452	0.996
P_0	0.95	0.746	0.712	0.157	0.672	0.768	0.849	0.767	0.114	0.531	0.892	0.941	0.799	0.082	0.386	0.996
<i>T = 300, Concentration Probability with $c = 0.03$</i>																
P_0	0.5	0.942	0.948	0.908	0.941	0.950	0.949	0.953	0.845	0.903	0.958	0.979	0.977	0.777	0.837	0.997
	0.7	0.902	0.908	0.770	0.892	0.915	0.929	0.932	0.681	0.826	0.945	0.976	0.973	0.602	0.747	0.999
	0.9	0.837	0.838	0.461	0.798	0.852	0.900	0.875	0.355	0.697	0.927	0.971	0.916	0.290	0.551	1.000
P_0	0.95	0.787	0.759	0.312	0.729	0.808	0.864	0.790	0.252	0.594	0.905	0.950	0.818	0.199	0.460	1.000

Table 2.5: The coverage and length of nominal 0.95-level confidence sets with small breaks $(\mu_{b,0}, \beta_{b,0}) = (0.5, 0.05)$.

$(\mu_{b,0}, \beta_{b,0}) = (0.5, 0.05)$																$\theta_0 = 0.3$								$\theta_0 = 0.6$								$\theta_0 = 0.9$							
		C_{ini}	C_{tr}^{**}	$C_{tr}^{provision}$	C_{tr}	$C_{tr}^{provision}$	C_{wld}	$C_{wld,provision}$	C_{bi}	C_{tr}^{**}	$C_{tr}^{provision}$	C_{tr}	$C_{tr}^{provision}$	C_{wld}	$C_{wld,provision}$	C_{bi}	C_{tr}^{**}	$C_{tr}^{provision}$	C_{tr}	$C_{tr}^{provision}$	C_{wld}	$C_{wld,provision}$	C_{bi}	C_{tr}^{**}	$C_{tr}^{provision}$	C_{tr}	$C_{tr}^{provision}$	C_{wld}	$C_{wld,provision}$										
$T = 150$, Coverage																																							
ρ_0	0.5	0.989	0.922	0.946	0.865	0.887	0.683	0.671	0.987	0.910	0.933	0.848	0.865	0.709	0.684	0.982	0.928	0.942	0.879	0.892	0.753	0.734																	
	0.7	0.981	0.892	0.923	0.780	0.805	0.671	0.670	0.979	0.899	0.929	0.795	0.819	0.704	0.691	0.967	0.918	0.944	0.837	0.874	0.731	0.720																	
	0.9	0.958	0.828	0.899	0.629	0.692	0.690	0.718	0.935	0.851	0.928	0.691	0.780	0.688	0.717	0.928	0.857	0.942	0.736	0.854	0.685	0.713																	
	0.95	0.925	0.785	0.887	0.589	0.665	0.680	0.723	0.901	0.790	0.912	0.630	0.745	0.676	0.713	0.897	0.808	0.931	0.675	0.841	0.654	0.706																	
$T = 150$, Length																																							
ρ_0	0.5	0.445	0.246	0.300	0.158	0.185	0.281	0.285	0.617	0.302	0.377	0.181	0.219	0.334	0.338	0.681	0.361	0.425	0.223	0.269	0.368	0.373																	
	0.7	0.731	0.399	0.484	0.205	0.246	0.372	0.386	0.807	0.479	0.569	0.252	0.308	0.419	0.437	0.829	0.537	0.620	0.319	0.402	0.451	0.465																	
	0.9	0.889	0.585	0.680	0.266	0.309	0.531	0.565	0.880	0.636	0.744	0.347	0.405	0.552	0.592	0.887	0.667	0.778	0.418	0.510	0.563	0.601																	
	0.95	0.870	0.581	0.694	0.278	0.311	0.555	0.605	0.857	0.619	0.747	0.348	0.396	0.569	0.623	0.861	0.643	0.770	0.411	0.504	0.565	0.612																	
$T = 300$, Coverage																																							
ρ_0	0.5	0.995	0.951	0.968	0.915	0.930	0.748	0.731	0.994	0.944	0.959	0.904	0.915	0.778	0.766	0.988	0.963	0.970	0.942	0.951	0.868	0.871																	
	0.7	0.994	0.933	0.954	0.869	0.883	0.767	0.750	0.989	0.946	0.954	0.891	0.896	0.802	0.798	0.984	0.965	0.970	0.938	0.947	0.891	0.885																	
	0.9	0.977	0.884	0.912	0.734	0.729	0.752	0.750	0.969	0.902	0.934	0.778	0.786	0.775	0.761	0.962	0.926	0.960	0.855	0.909	0.782	0.777																	
	0.95	0.952	0.834	0.902	0.622	0.651	0.715	0.735	0.941	0.858	0.930	0.688	0.745	0.716	0.737	0.937	0.881	0.956	0.768	0.881	0.714	0.756																	
$T = 300$, Length																																							
ρ_0	0.5	0.190	0.099	0.113	0.072	0.080	0.160	0.161	0.219	0.110	0.124	0.074	0.082	0.186	0.188	0.233	0.109	0.119	0.076	0.083	0.200	0.204																	
	0.7	0.269	0.137	0.170	0.087	0.102	0.222	0.230	0.296	0.157	0.198	0.098	0.117	0.242	0.255	0.349	0.165	0.208	0.109	0.136	0.268	0.282																	
	0.9	0.634	0.343	0.452	0.145	0.164	0.370	0.398	0.736	0.440	0.574	0.190	0.220	0.416	0.452	0.790	0.497	0.633	0.260	0.338	0.445	0.480																	
	0.95	0.776	0.488	0.604	0.174	0.187	0.458	0.499	0.827	0.563	0.683	0.235	0.244	0.499	0.544	0.853	0.605	0.728	0.322	0.386	0.521	0.569																	

Table 2.6: The coverage and length of nominal 0.95-level confidence sets with large breaks $(\mu_{b,0}, \beta_{b,0}) = (1, 0, 1)$.

$(\mu_{b,0}, \beta_{b,0}) = (1, 0.1)$		$\theta_0 = 0.3$												$\theta_0 = 0.6$												$\theta_0 = 0.9$												
		C_{bl}	C_{tr}^+	$C_{tr}^{prebrien}$	C_{tr}	$C_{tr}^{prebrien}$	C_{null}	$C_{null}^{prebrien}$	C_{bl}	C_{tr}^+	$C_{tr}^{prebrien}$	C_{tr}	$C_{tr}^{prebrien}$	C_{null}	$C_{null}^{prebrien}$	C_{bl}	C_{tr}^+	$C_{tr}^{prebrien}$	C_{tr}	$C_{tr}^{prebrien}$	C_{null}	$C_{null}^{prebrien}$	C_{bl}	C_{tr}^+	$C_{tr}^{prebrien}$	C_{tr}	$C_{tr}^{prebrien}$	C_{null}	$C_{null}^{prebrien}$	C_{bl}	C_{tr}^+	$C_{tr}^{prebrien}$	C_{tr}	$C_{tr}^{prebrien}$	C_{null}	$C_{null}^{prebrien}$		
$T = 150$, Coverage																																						
ρ_0	0.5	0.994	0.979	0.984	0.976	0.981	0.940	0.935	0.995	0.981	0.985	0.978	0.982	0.943	0.941	0.990	0.985	0.988	0.982	0.986	0.953	0.958																
	0.7	0.994	0.973	0.980	0.967	0.974	0.926	0.927	0.988	0.977	0.983	0.973	0.980	0.934	0.938	0.983	0.983	0.986	0.974	0.982	0.938	0.951																
	0.9	0.983	0.947	0.968	0.928	0.955	0.888	0.911	0.968	0.948	0.982	0.928	0.973	0.885	0.917	0.954	0.945	0.985	0.920	0.977	0.879	0.933																
	0.95	0.962	0.913	0.963	0.889	0.944	0.856	0.893	0.940	0.912	0.974	0.887	0.964	0.850	0.908	0.928	0.898	0.977	0.867	0.969	0.832	0.920																
$T = 150$, Length																																						
ρ_0	0.5	0.238	0.059	0.071	0.038	0.044	0.196	0.200	0.285	0.063	0.074	0.038	0.045	0.237	0.244	0.314	0.066	0.074	0.041	0.046	0.258	0.267																
	0.7	0.351	0.090	0.112	0.053	0.062	0.274	0.288	0.422	0.100	0.126	0.056	0.066	0.303	0.325	0.496	0.112	0.134	0.063	0.071	0.333	0.360																
	0.9	0.725	0.227	0.262	0.098	0.099	0.439	0.481	0.750	0.256	0.295	0.115	0.113	0.486	0.542	0.794	0.298	0.305	0.145	0.128	0.530	0.596																
	0.95	0.776	0.295	0.328	0.126	0.115	0.518	0.574	0.798	0.325	0.337	0.152	0.125	0.563	0.636	0.816	0.361	0.340	0.187	0.139	0.587	0.668																
$T = 300$, Coverage																																						
ρ_0	0.5	0.998	0.984	0.988	0.980	0.983	0.955	0.950	0.996	0.987	0.989	0.985	0.987	0.959	0.959	0.993	0.992	0.994	0.991	0.993	0.986	0.987																
	0.7	0.998	0.981	0.985	0.978	0.980	0.945	0.944	0.994	0.988	0.989	0.986	0.988	0.967	0.968	0.991	0.994	0.995	0.994	0.995	0.990	0.992																
	0.9	0.992	0.972	0.982	0.963	0.973	0.953	0.958	0.986	0.978	0.989	0.969	0.985	0.961	0.973	0.982	0.983	0.993	0.977	0.992	0.988	0.988																
	0.95	0.984	0.950	0.972	0.938	0.960	0.921	0.939	0.975	0.955	0.985	0.944	0.978	0.925	0.955	0.960	0.957	0.992	0.946	0.990	0.921	0.978																
$T = 300$, Length																																						
ρ_0	0.5	0.131	0.027	0.031	0.017	0.018	0.104	0.105	0.146	0.027	0.029	0.016	0.017	0.121	0.122	0.152	0.018	0.019	0.011	0.011	0.130	0.132																
	0.7	0.171	0.037	0.043	0.022	0.025	0.144	0.148	0.179	0.032	0.037	0.019	0.021	0.156	0.163	0.198	0.023	0.024	0.014	0.014	0.173	0.183																
	0.9	0.306	0.073	0.098	0.039	0.042	0.245	0.267	0.377	0.077	0.100	0.038	0.037	0.282	0.312	0.452	0.078	0.077	0.042	0.027	0.311	0.347																
	0.95	0.487	0.131	0.172	0.057	0.054	0.329	0.366	0.590	0.150	0.175	0.063	0.047	0.382	0.428	0.656	0.165	0.129	0.074	0.036	0.421	0.478																

3

Fully Modified GLS with an Application to the Environmental Kuznets Curve

This chapter is based on the paper Lin and Reuvers (2020b).

Abstract

In this chapter, we develop the asymptotic theory of a Fully Modified Generalized Least Squares estimator for multivariate cointegrating polynomial regressions. Such regressions allow for deterministic trends, stochastic trends and integer powers of stochastic trends to enter the cointegrating relations. Our fully modified estimator incorporates: (1) the direct estimation of the inverse autocovariance matrix of the multidimensional errors, and (2) second order bias corrections. The resulting estimator has the intuitive interpretation of applying a weighted least squares objective function to filtered data series. Moreover, the required second order bias corrections are convenient byproducts of our approach and lead to standard asymptotic inference. We also study several multivariate KPSS-type of tests for the null of cointegration. A comprehensive simulation study shows good performance of the FM-GLS estimator and the related tests. As a practical illustration, we reinvestigate the Environmental Kuznets Curve (EKC) hypothesis for six early industrialized countries as in Wagner et al. (2020).

3.1 Introduction

In recent years, there has been an increasing interest in the theoretical properties and theoretical justifications of nonlinear cointegrating relations. For theoretical properties we refer to the textbook treatise by Wang (2015), the recent review article by Tjøstheim (2020), and the extensive references found in either of them. Theoretical justifications are in some cases refinements of existing economic theory, e.g. nonlinear cointegration among bond yields with different times to maturity due to yield-dependent risk premia as discussed in Breitung (2001), or nonlinear purchasing power parity due to transaction/transportation costs and trade barriers (e.g. Hong and Phillips (2010)). In other cases, economic theory postulates a nonlinear cointegrating relation from the outset. A popular example of the latter is the Environmental Kuznets curve described in Grossman and Krueger (1995).¹

There are three branches of literature on the estimation of such nonlinear cointegrating relations. First, the papers by Park and Phillips (1999) and Park and Phillips (2001) are concerned with nonlinear cointegration analysis of a parametric form. Second, there is a literature on nonparametric kernel estimation of nonlinear cointegrating equations, see for example Wang and Phillips (2009) or Li et al. (2020). The third approach is reminiscent of a nonparametric sieve estimation with power polynomial basis. That is, one estimates a cointegrating relation containing integer powers of integrated regressors. Wagner and Hong (2016) named this a cointegrating polynomial regression (CPR). The multivariate seemingly unrelated regressions extension is available in Wagner et al. (2020). Our model specification builds on this Seemingly Unrelated Cointegrating Polynomial Regression (SUCPR) setup.

We make two theoretical contributions to the literature on cointegrating polynomial regressions. First, we propose the Fully Modified Generalized Least

¹There is no direct reference to Kuznets in the original paper by Grossman and Krueger (1995). But their nonlinear relations between environmental indicators and per capita GDP do remind strongly of the inverted U-shaped between income inequality and economic growth proposed by Kuznets (1901-1985). The term ‘environmental Kuznets curve’ was used later.

Squares (FM-GLS) estimator. This estimator requires two main steps: (1) It employs the inverse covariance matrix of the $2nT$ -dimensional innovation vector, that is, the covariance matrix of the vector which stacks the n disturbances in the cointegrating equations and the n disturbances driving the $I(1)$ regressors over the time span T . The estimation of this inverse covariance matrix is based on the Modified Cholesky Decomposition (MCD) originating from Pourahmadi (1999). The approach is computationally simple because the required quantities are obtained from the coefficients and prediction error variances of best linear least squares predictors. In our setting this translates into estimating multiple VAR models up to some maximum lag order q . Sufficient conditions for consistency are provided. (2) We exploit the previous results to correct the second-order biases, resulting in improved efficiency and standard chi-square inference. Also note that the approach differs from the linear cointegration results in Mark et al. (2005) and Moon and Perron (2005) since our bias corrections do not rely on leads and lags augmentation. Second, a multi-equation cointegration specification asks for a multivariate cointegration test. Building upon the work by Choi and Saikkonen (2010), we propose three such tests. The first test uses pre-filtered residuals to account for serial correlation, whereas the other two are direct multivariate generalizations of the KPSS-type of test in Wagner and Hong (2016). The estimator and cointegration tests are subsequently studied by Monte Carlo simulation. In our simulations, the FM-GLS estimator has a higher estimation accuracy and its implied Wald test has better size control and higher size-adjusted power. We find by simulation that prefiltering improves the size control of the cointegration tests but has an adverse effect on power. In the empirical application there is a surprisingly large spread in the widths of the confidence intervals. It turns out that FM-SUR, and to a lesser degree FM-SOLS, underestimates the parameter uncertainty compared to FM-GLS.

The plan of this chapter is as follows. Section 3.2 introduces the model and the modified Cholesky block decomposition. This decomposition is the main ingredient for the fully modified GLS estimator. The related asymptotic theory and stationarity tests are discussed in Section 3.3 whereas a finite sample

simulation study is presented in Section 3.4. The empirical application can be found in Section 3.5 where we look at the environmental Kuznets curve. Section 3.6 concludes. All proofs are collected in the Appendices.

Some words on notation. C denotes a generic positive constant. The integer part of the number $a \in \mathbb{R}^+$ is denoted by $[a]$. For a vector $\mathbf{x} \in \mathbb{R}^n$, its dimension is abbreviated by $\dim(\mathbf{x})$ and its p -norm by $\|\mathbf{x}\|_p = (\sum_{i=1}^n |x_i|^p)^{1/p}$. When applied to a matrix, $\|\mathbf{A}\|_p$ signifies the induced norm defined by $\|\mathbf{A}\|_p = \sup_{\mathbf{x} \neq \mathbf{0}} \|\mathbf{A}\mathbf{x}\|_p / \|\mathbf{x}\|_p$. The subscripts are omitted whenever $p = 2$, e.g. $\|\mathbf{x}\| = (\sum_{i=1}^n |x_i|^2)^{1/2}$ and $\|\mathbf{A}\| = (\lambda_{\max}(\mathbf{A}'\mathbf{A}))^{1/2}$ where $\lambda_{\max}(\cdot)$ is the largest eigenvalue. Similarly, $\lambda_{\min}(\cdot)$ denotes the smallest eigenvalue. The Frobenius norm is denoted as $\|\cdot\|_{\mathcal{F}}$. The $(n \times n)$ identity matrix is written as \mathbf{I}_n . The i^{th} row or i^{th} column of an arbitrary matrix \mathbf{A} are selected using $\text{col}_i(\mathbf{A})$ and $\text{row}_i(\mathbf{A})$, respectively. The Kronecker product is denoted “ \otimes ”. We use the symbol “ \Rightarrow ” to signify weak convergence and the symbol “ $\stackrel{d}{=}$ ” for equality in distribution. The stochastic order and strict stochastic order relations are indicated by $O_p(\cdot)$ and $o_p(\cdot)$.

3.2 The Model

As in Wagner et al. (2020), we study a system of seemingly unrelated cointegrating polynomial regressions (SUCPR), that is

$$\mathbf{y}_t = \mathbf{Z}_t' \boldsymbol{\beta} + \mathbf{u}_t, \quad \text{for } t = 1, 2, \dots, T, \quad (3.1)$$

where the dependent variable $\mathbf{y}_t := [y_{1t}, y_{2t}, \dots, y_{nt}]'$ and innovations $\mathbf{u}_t := [u_{1t}, u_{2t}, \dots, u_{nt}]'$ are $(n \times 1)$ random vectors. For the cross-sectional unit i , we use as explanatory variables: (1) deterministic components such as an intercept and polynomial time trends up to order d_i , and (2) integer powers of the $I(1)$ regressors x_{it} up to degree s_i . Defining $\mathbf{d}_{it} = [1, t, \dots, t^{d_i}]'$, $\mathbf{s}_{it} = [x_{it}, \dots, x_{it}^{s_i}]$, and $\mathbf{z}_{it} = [\mathbf{d}_{it}', \mathbf{s}_{it}']'$, we subsequently collect all explanatory variables in the block diagonal matrix $\mathbf{Z}_t = \text{diag}[\mathbf{z}_{1t}, \dots, \mathbf{z}_{nt}]$. We are interested in the d -dimensional parameter vector $\boldsymbol{\beta}$ where $d = \sum_{i=1}^n (d_i + s_i + 1)$. Overall, each

cross-sectional unit in (3.1) specifies a single cointegrating relation containing polynomials in deterministic and stochastic trends. For each i , the highest orders of these polynomials, i.e. d_i and s_i , are assumed to be fixed and known. We do not allow for cointegration in the cross-sectional dimension.

The innovation series $\{u_t\}$ is allowed to exhibit dependencies over time and across series. We assume that these dependencies can be modeled by a stationary VAR(∞) process, that is

$$\mathcal{A}(L)u_t = \left(I_n - \sum_{j=1}^{\infty} A_j L^j \right) u_t = \eta_t, \quad (3.2)$$

(see Assumption 3.1 for further details). Efficient estimation of the parameter vector β now requires the use of generalized least squares (GLS). Our Zellner (1962)-type GLS estimator relies on the inverse of the $(nT \times nT)$ matrix $\Sigma_u = \mathbb{E}(uu')$ where $u = [u'_1, u'_2, \dots, u'_T]'$. In this chapter, we directly estimate Σ_u^{-1} using a multivariate extension of the modified Cholesky decomposition by Pourahmadi (1999). This extension was named the *Modified Cholesky Block Decomposition* (MCBD) by Kim and Zimmerman (2012) and Kohli et al. (2016). The latter papers used the MCBD to parametrize the covariance matrix of multivariate longitudinal data. As in Beutner et al. (2021), we use the MCBD for the time series application mentioned above, i.e. the computation of Σ_u^{-1} . The decomposition is closely related to linear minimum MSE predictors.

We define

$$\begin{aligned} A(\ell) &= \begin{bmatrix} A_1(\ell) & \cdots & A_\ell(\ell) \end{bmatrix} \\ &= \arg \min_{(\Theta_1, \dots, \Theta_\ell) \in \mathbb{R}^{n \times n\ell}} \mathbb{E} \|u_t - \Theta_1 u_{t-1} - \cdots - \Theta_\ell u_{t-\ell}\|^2, \end{aligned} \quad (3.3)$$

and

$$\begin{aligned} S(\ell) &= \mathbb{E} [u_t - A_1(\ell)u_{t-1} - \cdots - A_\ell(\ell)u_{t-\ell}] \\ &\quad \times [u_t - A_1(\ell)u_{t-1} - \cdots - A_\ell(\ell)u_{t-\ell}]', \end{aligned} \quad (3.4)$$

where $S(0) = \mathbb{E}(\mathbf{u}_t \mathbf{u}_t')$. The inverse of $\Sigma_{\mathbf{u}}$ is then given by

$$\Sigma_{\mathbf{u}}^{-1} = \mathcal{M}'_{\mathbf{u}} \mathcal{S}_{\mathbf{u}}^{-1} \mathcal{M}_{\mathbf{u}}, \quad (3.5)$$

where $\mathcal{S}_{\mathbf{u}} = \text{diag}(S(0), S(1), \dots, S(T-1))$, $\mathcal{M}_{\mathbf{u}} = [\mathbf{m}_{\mathbf{u}}^{ij}]_{1 \leq i, j \leq T}$ with

$$\mathbf{m}_{\mathbf{u}}^{ij} = \begin{cases} \mathbf{0}_{n \times n}, & i < j, \\ \mathbf{I}_n, & i = j, \\ -\mathbf{A}_{i-j}(i-1), & 2 \leq i \leq T, 1 \leq j \leq i-1, \end{cases} \quad (3.6)$$

and the $\mathbf{A}_j(i)$ follow from the partitioning $\mathbf{A}(\ell) = [\mathbf{A}_1(\ell), \dots, \mathbf{A}_{\ell}(\ell)]$.

Weak stationarity of $\{\mathbf{u}_t\}$ implies that the block elements of $\mathcal{M}_{\mathbf{u}}$ being far below the main diagonal are small. This suggests a banding approach in which small elements are replaced by zeros. More specifically, we construct a *Banded Inverse Autocovariance Matrix* (BIAM) as

$$\Sigma_{\mathbf{u}}^{-1}(q) = \mathcal{M}'_{\mathbf{u}}(q) \mathcal{S}_{\mathbf{u}}^{-1}(q) \mathcal{M}_{\mathbf{u}}(q), \quad (3.7)$$

where $q \ll T$ is called the banding parameter, $\mathcal{S}_{\mathbf{u}}(q)$ and $\mathcal{M}_{\mathbf{u}}(q)$ are block matrices given by $\mathcal{S}_{\mathbf{u}}(q) = \text{diag}(S(0), S(1), \dots, S(q), \dots, S(q))$ and $\mathcal{M}_{\mathbf{u}}(q) = [\mathbf{m}_{\mathbf{u}}^{ij}(q)]_{1 \leq i, j \leq T}$ with

$$\mathbf{m}_{\mathbf{u}}^{ij}(q) = \begin{cases} \mathbf{0}_{n \times n}, & i < j \text{ or } \{q+1 < i \leq T, 1 \leq j \leq i-q-1\} \\ \mathbf{I}_n, & i = j \\ -\mathbf{A}_{i-j}(i-1), & 2 \leq i \leq q, 1 \leq j \leq i-1 \\ -\mathbf{A}_{i-j}(q), & q+1 \leq i \leq T, i-q \leq j \leq i-1. \end{cases} \quad (3.8)$$

Example 3.1

Consider a stationary n -dimensional VAR(3) process specified as $\mathbf{u}_t = \sum_{j=1}^3 \mathbf{A}_j \mathbf{u}_{t-j} + \boldsymbol{\eta}_t$ with $\boldsymbol{\eta}_t \stackrel{i.i.d.}{\sim} (\mathbf{0}, \Sigma_{\boldsymbol{\eta}})$. For $T = 4$, the MCB

$\Sigma_u^{-1} = \mathbf{M}'_u \mathbf{S}_u^{-1} \mathbf{M}_u$ is based on

$$\mathbf{M}_u = \begin{bmatrix} \mathbf{I}_n & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ -\mathbf{A}_1(1) & \mathbf{I}_n & \mathbf{0} & \mathbf{0} \\ -\mathbf{A}_2(2) & -\mathbf{A}_1(2) & \mathbf{I}_n & \mathbf{0} \\ -\mathbf{A}_3 & -\mathbf{A}_2 & -\mathbf{A}_1 & \mathbf{I}_n \end{bmatrix}, \quad \mathbf{S}_u = \begin{bmatrix} S^{(0)} & & & \\ & S^{(1)} & & \\ & & S^{(2)} & \\ & & & \Sigma_{\eta\eta} \end{bmatrix}. \quad (3.9)$$

Alternatively, with banding parameter $q = 2$, the related banded inverse autocovariance matrix is $\Sigma_u^{-1}(2) = \mathbf{M}'_u(2) \mathbf{S}_u^{-1}(2) \mathbf{M}_u(2)$ with

$$\mathbf{M}_u(2) = \begin{bmatrix} \mathbf{I}_n & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ -\mathbf{A}_1(1) & \mathbf{I}_n & \mathbf{0} & \mathbf{0} \\ -\mathbf{A}_2(2) & -\mathbf{A}_1(2) & \mathbf{I}_n & \mathbf{0} \\ \mathbf{0} & -\mathbf{A}_2(2) & -\mathbf{A}_1(2) & \mathbf{I}_n \end{bmatrix}, \quad \mathbf{S}_u(2) = \begin{bmatrix} S^{(0)} & & & \\ & S^{(1)} & & \\ & & S^{(2)} & \\ & & & S^{(2)} \end{bmatrix}. \quad (3.10)$$

The model of (3.1) can be stacked over time to yield the representation $\mathbf{y} = \mathbf{Z}\beta + \mathbf{u}$ with $\mathbf{y} = [\mathbf{y}'_1, \mathbf{y}'_2, \dots, \mathbf{y}'_T]'$, $\mathbf{Z} = [\mathbf{Z}_1, \mathbf{Z}_2, \dots, \mathbf{Z}_T]'$ and \mathbf{u} as before. For the moment, we assume $\Sigma_u^{-1}(q)$ is *known* and focus on the following estimator:

$$\widehat{\beta}_{GLS} := \left(\mathbf{Z}' \Sigma_u^{-1}(q) \mathbf{Z} \right)^{-1} \mathbf{Z}' \Sigma_u^{-1}(q) \mathbf{y}. \quad (3.11)$$

A discussion on the properties of this infeasible estimator is informative because: (1) the incurred estimation error of an appropriately constructed estimator $\widehat{\Sigma}_u^{-1}(q)$ will be asymptotically negligible, and (2) we can suppress the effect of banding by letting q increase with sample size.

Two remarks related to $\widehat{\beta}_{GLS}$ are instructive. First, the GLS estimator differs from the usual least squares estimator $\widehat{\beta}_{OLS} := (\mathbf{Z}'\mathbf{Z})^{-1} \mathbf{Z}'\mathbf{y}$ by a weighing with the inverse covariance matrix $\Sigma_u^{-1}(q)$. It is well documented in standard econometric textbooks (e.g. chapter 7 of Davidson and MacKinnon (2004)) that this weighing may lead to substantial efficiency gains. Second, it is illustrative to substitute the Modified Cholesky Decomposition of $\Sigma_u^{-1}(q)$ into the definition of this infeasible GLS estimator. The result is $\widehat{\beta}_{GLS} = (\mathbf{Z}'_{filt} \mathbf{S}_u^{-1}(q) \mathbf{Z}_{filt})^{-1} \mathbf{Z}'_{filt} \mathbf{S}_u^{-1}(q) \mathbf{y}_{filt}$ where $\mathbf{Z}_{filt} = \mathbf{M}_u(q) \mathbf{Z}$, and $\mathbf{y}_{filt} = \mathbf{M}_u(q) \mathbf{y}$. The premultiplications by $\mathbf{M}_u(q)$ have the effect of filtering and take care of serial correlation. $\mathbf{S}_u^{-1}(q)$ applies scaling and rotation to account for the correlations between the series. The following univariate autoregressive setting exemplifies this intuition.

Example 3.2

A linear trend model $y_t = \beta t + u_t$ has AR(1) innovations $u_t = \rho u_{t-1} + \eta_t$ where $\eta_t \stackrel{i.i.d.}{\sim} (0, \sigma^2)$ and $|\rho| < 1$. Taking $n = 1$, the expressions of Example 3.1 are easily adapted to yield: $\mathbf{S}_u = \text{diag}\left(\frac{\sigma^2}{1-\rho^2}, \sigma^2, \dots, \sigma^2\right)$,

$$\mathbf{M}_u \mathbf{y} = \begin{bmatrix} 1 & & & \\ -\rho & 1 & & \\ \vdots & \ddots & \ddots & \\ 0 & \cdots & -\rho & 1 \end{bmatrix} \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_T \end{bmatrix} = \begin{bmatrix} y_1 \\ y_2 - \rho y_1 \\ \vdots \\ y_T - \rho y_{T-1} \end{bmatrix},$$

and a similar transformation for the linear trend. The implied GLS estimator coincides with the estimator from Prais and Winsten (1954).

3.3 Asymptotic Theory

In this section, we present the asymptotic results. More specifically, we derive: (1) the limiting distribution of the GLS estimator, (2) the fully modified GLS (FM-GLS) estimator that corrects for second order bias terms, (3) a Wald test statistic, and (4) several multivariate KPSS-type of tests for the null of cointegration. We will also compare this FM-GLS estimator with the two fully modified estimators defined in Proposition 1 of Wagner et al. (2020). The following assumption will facilitate the development of the asymptotic theory.

Assumption 3.1 (Innovation Processes)

The innovations processes in the model satisfy the following assumptions:

- (a) The process $\zeta_t = [\eta'_t, \varepsilon'_t]'$ is an independent and identically distributed (i.i.d.) sequence with $\mathbb{E}(\zeta_t \zeta'_t) = \begin{bmatrix} \Sigma_{\eta\eta} & \Sigma_{\eta\varepsilon} \\ \Sigma_{\varepsilon\eta} & \Sigma_{\varepsilon\varepsilon} \end{bmatrix} > 0$ and $\mathbb{E}(\|\zeta_t\|^{2r}) \leq C_r < \infty$ for some constant $C_r > 0$ and some $r > 2$.
- (b) $\det(\mathcal{A}(z)) \neq 0$ for all $|z| \leq 1$ and $\sum_{j=0}^{\infty} j \|\mathbf{A}_j\|_{\mathcal{F}} < \infty$.

(c) $\Delta \mathbf{x}_t = \mathbf{v}_t$ admits the $\text{VAR}(\infty)$ process $\mathcal{D}(L)\mathbf{v}_t = \boldsymbol{\varepsilon}_t$, where $\mathcal{D}(L) = \mathbf{I}_n - \sum_{j=1}^{\infty} \mathbf{D}_j L^j$. Moreover, $\det(\mathcal{D}(z)) \neq 0$ for all $|z| \leq 1$ and $\sum_{j=0}^{\infty} j \|\mathbf{D}_j\|_{\mathcal{F}} < \infty$.

The stationary $\text{VAR}(\infty)$ specifications for $\{\mathbf{u}_t\}$ and $\{\mathbf{v}_t\}$ are natural given the linear minimum MSE predictor formulae that underly the definitions of the MCBD and BIAM. Moreover, the conditions in Assumption 3.1 ensure that the lag polynomials $\mathcal{A}(L)$ and $\mathcal{D}(L)$ are invertible (see for example Theorem 7.4.2 of Hannan and Deistler (2012)), thereby showing that our Assumption 3.1 is similar to the linear processes assumptions that are regularly adopted in the literature on nonlinear cointegration, cf. Choi and Saikkonen (2010), Wagner and Hong (2016), and Wagner et al. (2020). The assumption $\det(\mathcal{D}(1)) \neq 0$ rules out cointegration among the components of $\{\mathbf{x}_t\}$.

Under Assumption 3.1(a), an invariance principle holds for $\boldsymbol{\zeta}_t$, i.e. $\frac{1}{T^{1/2}} \sum_{t=1}^{\lfloor rT \rfloor} \boldsymbol{\zeta}_t \Rightarrow \mathbf{B}_{\boldsymbol{\zeta}}(r) \equiv \begin{bmatrix} \mathbf{B}_{\boldsymbol{\eta}}(r) \\ \mathbf{B}_{\boldsymbol{\epsilon}}(r) \end{bmatrix}$ where $\mathbf{B}_{\boldsymbol{\zeta}}$ denotes an $2n$ -dimensional Brownian motion with covariance matrix $\begin{bmatrix} \Sigma_{\boldsymbol{\eta}\boldsymbol{\eta}} & \Sigma_{\boldsymbol{\eta}\boldsymbol{\epsilon}} \\ \Sigma_{\boldsymbol{\epsilon}\boldsymbol{\eta}} & \Sigma_{\boldsymbol{\epsilon}\boldsymbol{\epsilon}} \end{bmatrix}$. Moreover, Assumptions 3.1(b)-(c) justify the use of the Beveridge-Nelson decomposition (Phillips and Solo (1992)). A functional central limit theorem for linear processes is thus also applicable to $\boldsymbol{\xi}_t = [\mathbf{u}'_t, \mathbf{v}'_t]'$, that is

$$\frac{1}{T^{1/2}} \sum_{t=1}^{\lfloor rT \rfloor} \boldsymbol{\xi}_t \Rightarrow \mathbf{B}_{\boldsymbol{\xi}}(r) \equiv \begin{bmatrix} \mathbf{B}_{\mathbf{u}}(r) \\ \mathbf{B}_{\mathbf{v}}(r) \end{bmatrix} \equiv \begin{bmatrix} \mathcal{A}(1) & \mathbf{0} \\ \mathbf{0} & \mathcal{D}(1) \end{bmatrix}^{-1} \begin{bmatrix} \mathbf{B}_{\boldsymbol{\eta}}(r) \\ \mathbf{B}_{\boldsymbol{\epsilon}}(r) \end{bmatrix}, \quad (3.12)$$

where the Brownian motion $\mathbf{B}_{\boldsymbol{\xi}}(r)$ of dimension $2n$ has the covariance matrix $\boldsymbol{\Omega} = \sum_{h=-\infty}^{\infty} \mathbb{E}(\boldsymbol{\xi}_t \boldsymbol{\xi}'_{t+h})$ given by

$$\boldsymbol{\Omega} \triangleq \begin{bmatrix} \boldsymbol{\Omega}_{uu} & \boldsymbol{\Omega}_{uv} \\ \boldsymbol{\Omega}_{vu} & \boldsymbol{\Omega}_{vv} \end{bmatrix} = \begin{bmatrix} \mathcal{A}(1) & \mathbf{0} \\ \mathbf{0} & \mathcal{D}(1) \end{bmatrix}^{-1} \begin{bmatrix} \Sigma_{\boldsymbol{\eta}\boldsymbol{\eta}} & \Sigma_{\boldsymbol{\eta}\boldsymbol{\epsilon}} \\ \Sigma_{\boldsymbol{\epsilon}\boldsymbol{\eta}} & \Sigma_{\boldsymbol{\epsilon}\boldsymbol{\epsilon}} \end{bmatrix} \begin{bmatrix} \mathcal{A}(1)' & \mathbf{0} \\ \mathbf{0} & \mathcal{D}(1)' \end{bmatrix}^{-1}. \quad (3.13)$$

Apart from this long-run covariance matrix $\boldsymbol{\Omega}$, we also introduce the one-sided long-run covariance matrix $\boldsymbol{\Delta} = \begin{bmatrix} \boldsymbol{\Delta}_{uu} & \boldsymbol{\Delta}_{uv} \\ \boldsymbol{\Delta}_{vu} & \boldsymbol{\Delta}_{vv} \end{bmatrix} = \sum_{h=0}^{\infty} \mathbb{E}(\boldsymbol{\xi}_t \boldsymbol{\xi}'_{t+h})$. The Brownian motion defined by $\mathbf{B}_{u,v} = \mathbf{B}_u - \boldsymbol{\Omega}_{uv} \boldsymbol{\Omega}_{vv}^{-1} \mathbf{B}_v$ is by construction orthogonal to \mathbf{B}_v . Its $(n \times n)$ covariance matrix equals $\boldsymbol{\Omega}_{u,v} = \boldsymbol{\Omega}_{uu} - \boldsymbol{\Omega}_{uv} \boldsymbol{\Omega}_{vv}^{-1} \boldsymbol{\Omega}_{vu}$.

3.3.1 Infeasible GLS

We start our analysis assuming that the $(nT \times nT)$ covariance matrix $\Sigma_u(q)$ is a known quantity for each q . The modified Cholesky block decomposition of page 94 can now be used to derive the limiting distribution of this infeasible GLS estimator. A insightful exposition of our results requires further notation.

- (a) Introduce scaling matrices: $G_{d_i,T} := T^{-1/2} \text{diag}[1, T^{-1}, \dots, T^{-d_i}]$ for the time trends, and $G_{s_i,T} := T^{-1/2} \text{diag}[T^{-1/2}, T^{-1}, \dots, T^{-s_i/2}]$ for the stochastic trends. Moreover, we define $G_T := \text{diag}[G_{1,T}, \dots, G_{n,T}]$, where $G_{i,T} := \text{diag}[G_{d_i,T}, G_{s_i,T}]$.
- (b) Let $d_i(r) := [1, r, \dots, r^{d_i}]'$, $B_{s_i}(r) := [B_{v_i}(r), B_{v_i}^2(r), \dots, B_{v_i}^{s_i}(r)]'$ and $j_i(r) := [d_i(r)', B_{s_i}(r)']'$. Define $d \times n$ block-diagonal random matrix $J(r) := \text{diag}[j_1(r), \dots, j_n(r)]$.
- (c) $b_i = [0'_{d_i+1}, 1, 2 \int_0^1 B_{v_i}(r)dr, \dots, s_i \int_0^1 B_{v_i}^{s_i-1}(r)dr]'$.

Finally, we use B_{v_j} as shorthand notation for the j^{th} component of B_v .

Theorem 3.1 (Limiting Distribution of the infeasible GLS Estimator)

If Assumption 3.1 holds, and if $q = q(T)$ satisfies $\frac{1}{q} + \frac{q}{T} \rightarrow 0$ as $T \rightarrow \infty$, then

$$\begin{aligned} G_T^{-1} (\widehat{\beta}_{GLS} - \beta) &\Rightarrow \left(\int_0^1 J(r) \Omega_{uu}^{-1} J(r)' dr \right)^{-1} \\ &\quad \times \left(\int_0^1 J(r) \Omega_{uu}^{-1} dB_{u,v}(r) + \int_0^1 J(r) \Omega_{uu}^{-1} \Omega_{uv} \Omega_{vv}^{-1} dB_v(r) + \mathcal{B} \right), \end{aligned} \quad (3.14)$$

where $\mathcal{B} = [\mathcal{B}'_1, \dots, \mathcal{B}'_n]'$ and $\mathcal{B}_i = \text{row}_i(\Sigma_{\epsilon\eta}) \text{col}_i(\Sigma_{\eta\eta}^{-1}) b_i$.

The limiting result in (3.14) coincides with the limiting distribution of the MSUR estimator, $\widetilde{\beta}_{MSUR} := (Z'(I_T \otimes \widehat{\Omega}_{uu}^{-1})Z)^{-1} (Z'(I_T \otimes \widehat{\Omega}_{uu}^{-1})y)$, as reported in Wagner et al. (2020), see their Proof of Proposition 1. The equivalence of these limiting distributions is caused by the facts that: (1) applying a linear filter to an integrated series only affect its long-run variance (e.g.

Phillips and Park (1988)), and (2) the previous statement remaining true when applying a linear filter to higher integer powers of integrated series. The terms $\int_0^1 \mathbf{J}(r) \boldsymbol{\Omega}_{uu}^{-1} \boldsymbol{\Omega}_{uv} \boldsymbol{\Omega}_{vv}^{-1} d\mathbf{B}_v(r)$ and \mathcal{B} in (3.14) reflect the presence of second order bias terms caused by serial correlation and endogeneity. In Section 3.3.3, we introduce the fully modified (FM) correction that adjust these bias terms and leads to standard inference. We first introduce a feasible version of the GLS estimator.

3.3.2 Consistent Estimation of $\boldsymbol{\Sigma}_u^{-1}(q)$ and Feasible GLS

Up to this point, we have discussed the infeasible estimator $\widehat{\boldsymbol{\beta}}_{GLS}$. A feasible GLS approach requires a consistent estimator of the $(nT \times nT)$ matrix $\boldsymbol{\Sigma}_u^{-1}(q)$. Several authors, e.g. Wu and Pourahmadi (2009) and McMurry and Politis (2010), have constructed consistent estimators of large covariance matrices using banding or tapering to reduce the number of unknown parameters. Direct usage of their results poses two difficulties because: (1) numerical inversion of large matrices is computationally expensive for large nT , and (2) matrix inversion might even be impossible because the estimated covariance matrix cannot be guaranteed to be positive definite. In the light of the such considerations, we will estimate $\boldsymbol{\Sigma}_u^{-1}(q)$ directly and ensure it to be positive definite. The approach is the sample counterpart of the BIAM described on page 95. That is, we replace true innovations by first stage OLS residuals $\widehat{\mathbf{u}}_t = \mathbf{y}_t - \mathbf{Z}_t \widehat{\boldsymbol{\beta}}_{OLS}$, and subsequently minimise a sample moment in estimated residuals rather than the population mean squared forecasting error. This method was previously used by Cheng et al. (2015) and Ing et al. (2016) for univariate time series. For a multivariate time series, we define

$$\begin{aligned} \widehat{\mathbf{A}}(\ell) &= \begin{bmatrix} \widehat{\mathbf{A}}_1(\ell) & \cdots & \widehat{\mathbf{A}}_\ell(\ell) \end{bmatrix} \\ &= \arg \min_{(\boldsymbol{\Theta}_1, \dots, \boldsymbol{\Theta}_\ell) \in \mathbb{R}^{n \times n\ell}} \sum_{t=\ell+1}^T \left\| \widehat{\mathbf{u}}_t - \boldsymbol{\Theta}_1 \widehat{\mathbf{u}}_{t-1} - \cdots - \boldsymbol{\Theta}_\ell \widehat{\mathbf{u}}_{t-\ell} \right\|^2 \end{aligned} \quad (3.15)$$

and

$$\widehat{S}(\ell) = \frac{1}{T-\ell} \sum_{t=\ell+1}^T \left[\widehat{u}_t - \widehat{A}_1(\ell)\widehat{u}_{t-1} - \cdots - \widehat{A}_\ell(\ell)\widehat{u}_{t-\ell} \right] \times \left[\widehat{u}_t - \widehat{A}_1(\ell)\widehat{u}_{t-1} - \cdots - \widehat{A}_\ell(\ell)\widehat{u}_{t-\ell} \right]', \quad (3.16)$$

where $\ell = 1, \dots, q$ and $\widehat{S}(0) = \frac{1}{T} \sum_{t=1}^T \widehat{u}_t \widehat{u}_t'$. Similarly to (3.7)-(3.8), we subsequently construct the matrices $\widehat{\mathcal{M}}_u(q) = [\widehat{m}_u^{ij}(q)]_{1 \leq i, j \leq T}$ and $\widehat{\mathcal{S}}_u(q) = \text{diag}(\widehat{S}(0), \widehat{S}(1), \dots, \widehat{S}(q), \dots, \widehat{S}(q))$, and obtain the implied multivariate BIAM estimator as

$$\widehat{\Sigma}_u^{-1}(q) = \widehat{\mathcal{M}}_u'(q) \widehat{\mathcal{S}}_u^{-1}(q) \widehat{\mathcal{M}}_u(q). \quad (3.17)$$

Assumption 3.2

For $\widehat{u} = [\widehat{u}_1', \dots, \widehat{u}_T']'$ and $u = [u_1', \dots, u_T']'$, assume $\|\widehat{u} - u\|^2 = O_p(1)$.

Assumption 3.3

Assume $q = q_T$ satisfies $\frac{1}{q_T} + \frac{q_T^3}{T} \rightarrow 0$ as $T \rightarrow \infty$.

Assumption 3.2 requires the residuals to be sufficiently close to the true innovations. It is a rather mild assumption and it is satisfied if residuals are computed by least squares. Assumption 3.3 places constraints on the banding parameter q_T . First, Assumption 3.3 requires the banding parameter to diverge with sample size. This ensures that no nonzero elements are (asymptotically) set to zero. Moreover, the assumption $q_T^3/T \rightarrow 0$ establishes an upper bound for the growth rate of q_T . The definition of $\widehat{A}(\ell)$, see (3.15), shows that we are fitting a vector autoregression (VAR) of increasing lag order to the residuals. Identical rate requirements are reported by Lewis and Reinsel (1985) when they derive consistency and asymptotic normality results when finite VAR models are fitted to infinite order VAR processes. The following theorem shows the consistent estimation of Σ_u^{-1} and implies that the infeasible and feasible GLS estimator have the same limiting distribution.

Theorem 3.2 (Consistent Estimation of Σ_u^{-1})

If Assumptions 3.1-3.3 hold, then as $T \rightarrow \infty$,

$$\begin{aligned} \left\| \widehat{\Sigma_u^{-1}}(q_T) - \Sigma_u^{-1} \right\| &\leq \left\| \widehat{\Sigma_u^{-1}}(q_T) - \Sigma_u^{-1}(q_T) \right\| + \left\| \Sigma_u^{-1}(q_T) - \Sigma_u^{-1} \right\| \\ &= O_p \left(\sqrt{q_T^3/T} \right) + O \left(\frac{1}{\sqrt{q_T}} \sum_{s=q_T+1}^{\infty} s \|A_s\|_{\mathcal{F}} \right) = o_p(1). \end{aligned} \quad (3.18)$$

3.3.3 Fully Modified Inference

The asymptotic results of Theorem 3.1 is not immediately useful for statistical inference. There are two difficulties. First, the second order bias dislocates the limiting distribution which can translate into substantial finite sample bias. This leads to a loss in efficiency. Second, possible dependencies between the Brownian motions B_u and B_v cause the limiting distribution to depend on nuisance parameters. Critical values would therefore be nuisance parameter dependent as well.

These two issues have received extensive attention in the linear cointegration literature. A (non-exhaustive) list of solution methods is: joint modeling as in Phillips (1991), Saikkonen's (1992) dynamic least squares, and the integrated modified OLS and fixed-b approaches by Vogelsang and Wagner (2014). We rely on the fully modified (FM) approach advocated by Phillips and Hansen (1990) and Phillips (1995). The idea is a twofold modification of the estimator: (1) second order bias terms are subtracted, and (2) a transformation of the dependent variable is introduced to obtain a zero-mean Gaussian mixture limiting distribution. Recently, Wagner et al. (2020) have proposed two estimators within the framework of seemingly unrelated cointegrating polynomial regressions. These estimators, FM-SOLS and FM-SUR, rely on kernel estimators of the one- and two-sided long-run covariance matrix (see Theorem 3.3). As such, we introduce the following assumption.

Assumption 3.4 (Consistent Estimation of Long-run Covariance Matrices)

$\widehat{\Omega}$ and $\widehat{\Delta}$ are consistent kernel estimators of the long-run covariance matrix Ω and the one-sided long-run covariance matrix Δ , respectively.

Andrews (1991) and Newey and West (1994) use kernel estimators for long-run covariance estimation. Their method involves the calculation of weighted sums of the autocovariance matrices of the residuals. These weights are determined by a kernel function and bandwidth parameter. Our Assumption 3.4 is easily satisfied by imposing suitable conditions on the kernel function and bandwidth parameter. We refer to Phillips (1995) and Jansson (2002) for an enumeration of such conditions.

Alternatively, we can obtain consistent one- and two-sided long-run covariance estimators within the BIAM framework of Section 3.3.2.² This approach resembles Berk (1974). The GLS estimator and its FM counterpart are thus constructed within a single framework. The estimators are as follows. For all $t = 1, 2, \dots, T$, we first stack \widehat{u}_t and $\Delta x_t = v_t$ in the $2n$ -dimensional vector $\widehat{\xi}_t = [\widehat{u}_t', \Delta x_t']'$. Since the BIAM estimator is fitting VAR processes up to order q_T , we will use the estimated $\text{VAR}(q_T)$ approximations to define the long-run covariance estimators. For Ω , the estimator is $\widehat{\Omega}_{q_T} = \left(\mathbf{I}_{2n} - \sum_{j=1}^{q_T} \widehat{F}_j(q_T) \right)^{-1} \widehat{\Sigma}_{q_T} \left(\mathbf{I}_{2n} - \sum_{j=1}^{q_T} \widehat{F}_j'(q_T) \right)^{-1}$, where $\widehat{\Sigma}_{q_T} = \widehat{S}(q_T)$ and $\widehat{F}_j(q_T)$ denote respectively the estimated prediction error variance and the coefficient matrix of the j^{th} lag when a $\text{VAR}(q_T)$ is fitted to $\{\widehat{\xi}_t\}_{t=1}^T$. The population one-sided long-run covariance matrix is $\Delta = \sum_{h=0}^{\infty} \mathbb{E}(\xi_t \xi_{t+h}')$. It is thus intuitive to approximate this quantity by a finite sum of estimated covariance matrices of $\{\widehat{\xi}_t\}_{t=1}^T$. These covariance matrices are nothing but subblocks of the matrix $\widehat{\Sigma}_{\xi}(q_T) = \widehat{\mathcal{M}}_{\xi}^{-1}(q_T) \widehat{S}_{\xi}(q_T) \widehat{\mathcal{M}}_{\xi}^{-1'}(q_T)$.³

²An overview of the procedure is given here. Appendix D provides further details.

³We use Σ_{ξ} to denote the $(2nT \times 2nT)$ matrix $\mathbb{E}(\xi \xi')$ where $\xi = [\xi_1', \xi_2', \dots, \xi_T']'$. The matrices $\widehat{\mathcal{M}}_{\xi}(q)$ and $\widehat{S}_{\xi}(q)$ are defined similarly to respectively $\widehat{\mathcal{M}}_u(q)$ and $\widehat{S}_u(q)$ (see page 101). The matrix $\widehat{\mathcal{M}}_{\xi}(q_T)$ is lower triangular with identity matrices on the main diagonal. Therefore, its matrix inverse exists and is fast to compute.

We therefore use

$$\widehat{\Delta}_{q_T, r_T} = \mathbf{Q}'_{r_T} \widehat{\Sigma}_{\xi}(q_T) \mathbf{Q}_1 \quad (3.19)$$

where $\mathbf{Q}_r = [\mathbf{O}_{2n \times 2n}, \dots, \mathbf{O}_{2n \times 2n}, \mathbf{I}_{2n}, \dots, \mathbf{I}_{2n}]'$ is an $(2nT \times 2n)$ block matrix of zeros of which the last r blocks have been replaced by identity matrices. To ensure consistency, we place the following rate restriction on the number of included autocovariance matrices.

Assumption 3.5

As $T \rightarrow \infty$, $r_T \rightarrow \infty$, $\frac{r_T q_T^3}{T} \rightarrow 0$, and $r_T = O(q_T)$.

Definitions and limiting results for FM estimators are presented in Theorem 3.3. The FM-SOLS, FM-SUR and FM-GLS estimator all depend on estimators for Δ and Ω . It is only the consistency of these estimators that is relevant for the asymptotic analysis, not whether the kernel or BIAM approach is employed. As such, we will not complicate notation by introducing additional notation to indicate whether the kernel or BIAM approach is used. In subsequent theorems, simulation results and the empirical application we will use kernel estimators for FM-SOLS and FM-SUR, and the BIAM approach for FM-GLS. This seems to be the logical choice for these estimators.

Theorem 3.3

For $i = 1, \dots, n$, define $\widehat{\mathbf{b}}_i = [\mathbf{0}'_{d_i+1}, T, 2 \sum_{t=1}^T x_{it}, \dots, s_i \sum_{t=1}^T x_{it}^{s_i-1}]'$, for $i = 1, \dots, n$. Also, define the $(n \times n)$ matrix $\widehat{\Delta}_{vu}^+$ as the (implied) consistent estimator of $\Delta_{vu}^+ = \Delta_{vu} - \Delta_{vv} \Omega_{vv}^{-1} \Omega_{vu}$.

(a) Define the FM-SOLS estimator as

$$\widehat{\beta}_{SOLS}^+ = (\mathbf{Z}' \mathbf{Z})^{-1} (\mathbf{Z}' \mathbf{y}^+ - \widehat{\mathbf{A}}), \quad (3.20)$$

where $\mathbf{y}^+ := [\mathbf{y}_1^+, \mathbf{y}_2^+, \dots, \mathbf{y}_n^+]'$ with $\mathbf{y}_i^+ = \mathbf{y}_i - \widehat{\Omega}_{uv} \widehat{\Omega}_{vv}^{-1} \Delta \mathbf{x}_i$, and $\widehat{\mathbf{A}} := [\widehat{\mathbf{A}}_1', \dots, \widehat{\mathbf{A}}_n']'$ with $\widehat{\mathbf{A}}_i = \widehat{\Delta}_{v_i u_i}^+ \widehat{\mathbf{b}}_i$ and $\widehat{\Delta}_{v_i u_i}^+$ being the i^{th} element on the

main diagonal of $\widehat{\Delta}_{vu}^+$. If Assumptions 3.1 and 3.4 hold, then

$$\mathbf{G}_T^{-1} (\widehat{\beta}_{SOLS}^+ - \beta) \Rightarrow \left(\int_0^1 \mathbf{J}(r) \mathbf{J}(r)' dr \right)^{-1} \int_0^1 \mathbf{J}(r) d\mathbf{B}_{u,v}(r). \quad (3.21)$$

(b) Define the FM-SUR estimator as

$$\widehat{\beta}_{SUR}^+ = \left(\mathbf{Z}' (\mathbf{I}_T \otimes \widehat{\Omega}_{u,v}^{-1}) \mathbf{Z} \right)^{-1} \left(\mathbf{Z}' (\mathbf{I}_T \otimes \widehat{\Omega}_{u,v}^{-1}) \mathbf{y}^+ - \widetilde{\mathbf{A}}^* \right), \quad (3.22)$$

where $\widetilde{\mathbf{A}}^* := [\widetilde{\mathbf{A}}_1^*, \dots, \widetilde{\mathbf{A}}_n^*]$ with $\widetilde{\mathbf{A}}_i^* = \text{row}_i(\widehat{\Delta}_{vu}^+) \text{col}_i(\widehat{\Omega}_{u,v}^{-1}) \widehat{\mathbf{b}}_i$. If Assumptions 3.1 and 3.4 hold, then

$$\mathbf{G}_T^{-1} (\widehat{\beta}_{SUR}^+ - \beta) \Rightarrow \left(\int_0^1 \mathbf{J}(r) \Omega_{u,v}^{-1} \mathbf{J}(r)' dr \right)^{-1} \int_0^1 \mathbf{J}(r) \Omega_{u,v}^{-1} d\mathbf{B}_{u,v}(r). \quad (3.23)$$

(c) Define the FM-GLS estimator as

$$\widehat{\beta}_{FGLS}^+ = \left(\mathbf{Z}' \widehat{\Sigma}_u^{-1}(q) \mathbf{Z} \right)^{-1} \left[\mathbf{Z}' \widehat{\Sigma}_u^{-1}(q) \mathbf{y} - \mathbf{Z}' (\mathbf{I}_T \otimes \widehat{\Omega}_{uu}^{-1} \widehat{\Omega}_{uv} \widehat{\Omega}_{vv}^{-1}) \mathbf{v} - \widehat{\mathcal{B}}^+ \right], \quad (3.24)$$

where $\mathbf{v} := [\Delta \mathbf{x}'_1, \dots, \Delta \mathbf{x}'_T]' = [\mathbf{v}'_1, \dots, \mathbf{v}'_T]'$, and $\widehat{\mathcal{B}}^+ = [\widehat{\mathcal{B}}_1^{+'}, \dots, \widehat{\mathcal{B}}_n^{+'}]'$ with

$$\widehat{\mathcal{B}}_i^+ = \left[\text{row}_i(\widehat{\Sigma}_{\epsilon\eta}) \text{col}_i(\widehat{\Sigma}_{\eta\eta}^{-1}) - \text{row}_i(\widehat{\Delta}_{vv}) \text{col}_i(\widehat{\Omega}_{vv}^{-1} \widehat{\Omega}_{vu} \widehat{\Omega}_{uu}^{-1}) \right] \widehat{\mathbf{b}}_i.$$

If Assumptions 3.1-3.3 and 3.5 hold, then

$$\mathbf{G}_T^{-1} (\widehat{\beta}_{FGLS}^+ - \beta) \Rightarrow \left(\int_0^1 \mathbf{J}(r) \Omega_{uu}^{-1} \mathbf{J}(r)' dr \right)^{-1} \int_0^1 \mathbf{J}(r) \Omega_{uu}^{-1} d\mathbf{B}_{u,v}(r). \quad (3.25)$$

The FM-GLS estimator is new to the seemingly unrelated CPR literature, whereas the FM-SOLS and FM-SUR estimators have recently appeared in Wagner et al. (2020). Theorem 3.3 indicates that all three estimators have a zero-mean Gaussian mixture limiting distribution implying that standard

inference is applicable for each. However, we also see from Theorem 3.3 that the limiting distributions are generally different because different types of weighing are used in the construction of the estimators.⁴

For completeness, we also detail how the FM-GLS estimator can be used to test linear hypotheses. A formal presentation of such a result is more involved because of the different convergence rates of the individual parameter estimators. That is, the parameters with the lowest convergence rate will dominate the asymptotic distribution and one should take care to avoid a degenerate limiting distribution. We will rule out such complications by considering hypothesis tests on individual parameters.⁵ Therefore, let \mathbf{R} denote a $(k \times s)$ selection matrix in which every row contains a single 1 and zeros otherwise. The null hypothesis $\mathbf{R}\beta = \mathbf{r}$ can be tested using the standard chi-squared limiting distribution of the Wald statistic (Theorem 3.4). These tests are practically relevant. For example, exclusion restrictions of the type $\mathbf{R}\beta = \mathbf{0}$ allow us to test whether the cointegrating relation is linear.

Theorem 3.4

Consider the null hypothesis $H_0 : \mathbf{R}\beta = \mathbf{r}$, which imposes k linearly independent restrictions. Under the assumptions of Theorem 3.3(c), the Wald test statistic

$$\mathcal{W} = (\mathbf{R}\widehat{\beta}_{FGLS}^+ - \mathbf{r})' \widehat{\Phi}^{-1} (\mathbf{R}\widehat{\beta}_{FGLS}^+ - \mathbf{r}) \Rightarrow \chi_k^2, \quad (3.26)$$

where

$$\widehat{\Phi} = \mathbf{R} \left[\mathbf{Z}' (\mathbf{I}_T \otimes \widehat{\Omega}_{uu}^{-1}) \mathbf{Z} \right]^{-1} \left[\mathbf{Z}' (\mathbf{I}_T \otimes \widehat{\Omega}_{uu}^{-1} \widehat{\Omega}_{u,v} \widehat{\Omega}_{uu}^{-1}) \mathbf{Z} \right] \left[\mathbf{Z}' (\mathbf{I}_T \otimes \widehat{\Omega}_{uu}^{-1}) \mathbf{Z} \right]^{-1} \mathbf{R}'.$$

⁴There are special cases in which some (pairs of) estimators become asymptotically equivalent. For example, if $n = 1$, then all estimators are asymptotically equivalent because the weighting matrices $\Omega_{u,v}^{-1}$ and Ω_{uu}^{-1} are now scalars. Also, under exogeneity, we have $\Omega_{uu} = \Omega_{u,v}$ and the FM-SUR and FM-GLS estimators share the same limiting distribution.

⁵For general linear hypothesis, we refer the reader to Sims et al. (1990) where a reordering based on convergence rates is used to establish the limiting distribution of the Wald F statistic for general linear hypothesis. The same approach is applicable in our setting but we will not explore this in greater detail.

3.3.4 Testing the Null of Cointegration

Stationarity tests are used to avoid spurious regressions and to verify the correct specification of the cointegrating relation. To test for stationarity of the seemingly unrelated cointegrating polynomial regressions (SUCPR) errors, we combine the test statistic from Nyblom and Harvey (2000) with the subsampling approach found in Choi and Saikkonen (2010) and Wagner and Hong (2016). We consider three test statistics. To treat all test statistics in a unified framework, we define

$$\varphi_{j,b}(\{x\}) = \left[x'_j, \sum_{s=j}^{j+1} x'_s, \dots, \sum_{s=j}^{j+b-1} x'_s \right]', \quad (3.27)$$

that is, a vector of length nb stacking the cumulative sums of $\{x_j, \dots, x_{j+b-1}\}$. If the innovations $\{u_t\}_{t=1}^T$ were known, we could use the full-sample KPSS-type of test $T^{-2}\varphi_{1,T}(\{u\})' (I_T \otimes \widehat{\Omega}_{uu}^{-1}) \varphi_{1,T}(\{u\}) = \text{tr} \left[\widehat{\Omega}_{uu}^{-1} T^{-2} \sum_{t=1}^T \left(\sum_{s=1}^t u_s \right) \left(\sum_{s=1}^t u_s \right)' \right]$ to test for stationarity of the innovations. Under the null of stationarity, this test statistic would converge weakly to $\int_0^1 \|W(r)\|^2 dr$ with $W(r)$ denoting an n -dimensional standard Brownian motion. This limiting distribution is free of nuisance parameters and the cumulative distribution function is available as a series expansion derived in Appendix C.

The innovations u_1, \dots, u_T are only available when cointegrating relations are pre-specified. If these coefficients are estimated, then this additional parameter uncertainty will contaminate the limiting distribution with nuisance parameters.⁶ The idea behind the subsampling approach is to construct a test statistic incorporating $b = b_T$ residuals while computing parameter estimators from all T observations. If b_T increases slowly with sample size, then the parameter estimation error will be negligible relative to the randomness in the errors and the asymptotic distribution remains $\int_0^1 \|W(r)\|^2 dr$.

⁶There are exceptions. Shin (1994) reports a nuisance parameter free limiting distribution for a single-equation linear cointegrating relation. This remains true if only a single integrated variable enters the cointegrating regression with a higher power, see Proposition 5 in Wagner and Hong (2016).

The three KPSS-type of test are based on the following residuals: $\hat{\mathbf{u}}_{i,SOLS}^+ = \mathbf{y}_t^+ - \mathbf{Z}_t \widehat{\boldsymbol{\beta}}_{SOLS}^+$, $\hat{\mathbf{u}}_{i,SUR}^+ = \mathbf{y}_t^+ - \mathbf{Z}_t \widehat{\boldsymbol{\beta}}_{SUR}^+$, and $\hat{\mathbf{u}}_{i,FGLS} = \mathbf{y}_t - \mathbf{Z}_t \widehat{\boldsymbol{\beta}}_{FGLS}^+$. The test statistic are:

$$K_{j,b_T}^i = b_T^{-2} \boldsymbol{\varphi}_{j,b_T}(\{\hat{\mathbf{u}}_i^+\})' \left(\mathbf{I}_{b_T} \otimes \widehat{\boldsymbol{\Omega}}_{u,v}^{-1} \right) \boldsymbol{\varphi}_{j,b_T}(\{\hat{\mathbf{u}}_i^+\}) \quad (3.28)$$

for $i \in \{SOLS, SUR\}$, and

$$K_{j,b_T}^{BIAM} = b_T^{-2} \boldsymbol{\varphi}_{j,b_T}(\{\hat{\mathbf{u}}_{FGLS}\})' \widehat{\boldsymbol{\Sigma}}_u^{-1}(q_T, b_T) \boldsymbol{\varphi}_{j,b_T}(\{\hat{\mathbf{u}}_{FGLS}\}), \quad (3.29)$$

where $\widehat{\boldsymbol{\Sigma}}_u^{-1}(q_T, b_T)$ is the $(nb_T \times nb_T)$ submatrix of $\widehat{\boldsymbol{\Sigma}}_u^{-1}(q_T)$ obtained by selecting the rows and columns related to all time indices in the set $\{n(T - b_T) + 1, n(T - b_T) + 2, \dots, nT\}$. The test statistic in (3.29) fits naturally into the FM-GLS estimation framework.

Theorem 3.5

Let the assumptions from Theorem 3.3 hold.

(a) If $\frac{1}{b_T} + \frac{b_T}{T} \rightarrow 0$ as $T \rightarrow \infty$, then

$$K_{j,b_T}^i \Rightarrow \int_0^1 \|\mathbf{W}(r)\|^2 dr, \quad 1 \leq j \leq T - b_T + 1, \quad \text{for } i \in \{SOLS, SUR\}.$$

(b) If $\frac{q_T}{b_T} + \frac{b_T}{T} \rightarrow 0$ as $T \rightarrow \infty$, then $K_{j,b_T}^{BIAM} \Rightarrow \int_0^1 \|\mathbf{W}(r)\|^2 dr$ for any $1 \leq j \leq T - b_T + 1$.

A sample of size T allows for up to $M = \lfloor T/b_T \rfloor$ series of nonoverlapping blocks of residuals of length b_T . Similarly to Choi and Saikkonen (2010), we apply the Bonferroni procedure to use all these series and thereby increase power. The approach is applicable to any of the three test statistics in Theorem 3.5. As such, we keep the notation general and use a generic K_j to denote a test statistic based on the j^{th} subseries, $j = 1, 2, \dots, M$. In the Bonferroni adjustment we compute $K_{max} = \{K_1, K_2, \dots, K_M\}$ and do

not reject the null hypothesis whenever $K_{max} \leq c_{\alpha/M}$ with $c_{\alpha/M}$ defined by $\mathbb{P}\left(\int_0^1 \|\mathbf{W}(r)\|^2 dr \geq c_{\alpha/M}\right) = \alpha/M$. The Bonferroni inequality implies $\lim_{T \rightarrow \infty} \mathbb{P}(K_{max} \leq c_{\alpha/M}) \geq 1 - \lim_{T \rightarrow \infty} \sum_{j=1}^M \mathbb{P}(K_j > c_{\alpha/M}) = 1 - \alpha$ and we see that the type-I error is under control by α asymptotically.

Remark 3.1

We suggest to follow Choi and Saikkonen (2010) in terms of the implementation of the subsampling approach. That is, the block size b_T is selected using the minimum volatility rule by Romano and Wolf (2001). For this particular block size we subsequently select subsamples by taking non-overlapping blocks from alternatively the start and the end of the sample.

Remark 3.2

The limiting results in Theorem 3.5 guarantee a correct asymptotic size. Our simulations show (1) that these tests have power against various alternative hypotheses and (2) that power increases with sample size. A theoretical investigation of the power properties is outside of the scope of this chapter.

3.4 Simulations

We now study the finite sample performance of the estimators and stationarity tests. First, we compare the FM-GLS estimator with the FM-SOLS and FM-SUR estimators from Wagner et al. (2020). All long-run covariance matrices are computed using a Bartlett kernel and the automatic bandwidth selection approach due to Andrews (1991). For FM-GLS, the banding parameter q_T is selected using the subsampling and risk-minimization approach explained in section 5 from Bickel and Levina (2008).⁷ Infeasible counterparts of the estimator are constructed assuming the knowledge of the true serial correlation and/or cross-sectional dependence pattern. These estimators are denoted by infSOLS, infSUR, and infGLS. Second, we look at the cointegration tests. We consider three test statistics: K^{SOLS} and K^{SUR} use the residuals as in (3.28),

⁷More details concerning the implementation can be found in appendix.

whereas K^{BIAM} employs the pre-filtered residuals from (3.29). All tests are implemented with minimum volatility block size selection and Bonferroni correction.

We consider $T \in \{100, 200, 500\}$ and $n \in \{3, 5\}$. All tests are performed at a nominal significant level of 5%. For stationary processes, a presample of 200 observations is used to remove the influence of the starting values. All results are based on 2.5×10^4 Monte Carlo replicates.

3.4.1 Monte Carlo Designs

We generate data according to a quadratic seemingly unrelated CPR. That is, we adopt the DGP in (3.1) with $z_{it} = [1, t, x_{it}, x_{it}^2]'$. The integrated variables satisfy $x_0 = \mathbf{0}$ and $\Delta x_t = v_t$. We explore two error processes.

Setting A (Errors as in Wagner et al. (2020)): As a benchmark, we revisit the simulation setting in Wagner et al. (2020) and generate innovations by

$$u_t = \rho_1 u_{t-1} + \varepsilon_t + \rho_2 e_t, \quad v_t = e_t + 0.5 e_{t-1}, \quad (3.30)$$

where $\varepsilon_t \stackrel{i.i.d.}{\sim} N(0, \Sigma(\rho_3))$, $e_t \stackrel{i.i.d.}{\sim} N(0, \Sigma(\rho_4))$ and

$$\Sigma(\rho) = \begin{bmatrix} 1 & \rho & \cdots & \rho \\ \rho & 1 & \cdots & \rho \\ \vdots & \vdots & \ddots & \vdots \\ \rho & \rho & \cdots & 1 \end{bmatrix} \quad (3.31)$$

is a symmetric Toeplitz matrix. The parameter ρ_1 controls the level of serial correlation and ρ_2 measures the degree of endogeneity. The parameters ρ_3 and ρ_4 indicate the extent of correlation across equations induced through ε_t and e_t , respectively. For simplicity, we assume identical values $\rho_1 = \rho_2 = \rho_3 = \rho_4 = \rho \in \{0, 0.3, 0.6, 0.8\}$. The true coefficient vector is $\beta = [\beta'_1, \dots, \beta'_n]'$, where $\beta_i = [1, 1, 5, \beta_{i,4}]'$ with $\beta_{i,4} = -0.3$, $i = 1, \dots, n$.

Setting B (VARMA Errors): To further investigate the importance of serial correlation, we consider a second specification of the innovation process:

$$\mathbf{u}_t = \mathbf{A}_1 \mathbf{u}_{t-1} + \boldsymbol{\eta}_t + \mathbf{A}_2 \boldsymbol{\eta}_{t-1}, \quad \mathbf{v}_t = \mathbf{A}_3 \mathbf{v}_{t-1} + \boldsymbol{\varepsilon}_t, \quad (3.32)$$

where $\boldsymbol{\eta}_t$ and $\boldsymbol{\varepsilon}_t$ are generated as $\begin{bmatrix} \boldsymbol{\eta}_t \\ \boldsymbol{\varepsilon}_t \end{bmatrix} \stackrel{i.i.d.}{\sim} \mathbf{N}(\mathbf{0}, \boldsymbol{\Sigma}(\theta))$ and $\boldsymbol{\Sigma}(\theta) \in \mathbb{R}^{2n \times 2n}$ as in (3.31) but with parameter θ . The matrices \mathbf{A}_i ($i = 1, 2, 3$) are generated independently and similarly to Chang et al. (2004). That is, we take the following three steps:

- (a) Generate an $n \times n$ random matrix \mathbf{U}_i from $\mathbf{U}[0, 1]$ and construct the orthogonal matrix $\mathbf{H}_i = \mathbf{U}_i (\mathbf{U}_i' \mathbf{U}_i)^{-1/2}$.
- (b) Generate n eigenvalues $\lambda_{i1}, \dots, \lambda_{in} \stackrel{i.i.d.}{\sim} \mathbf{U}[\underline{\lambda}, \bar{\lambda}]$.
- (c) Let $\mathbf{L}_i = \text{diag}(\lambda_{i1}, \dots, \lambda_{in})$ and compute $\mathbf{A}_i = \mathbf{H}_i \mathbf{L}_i \mathbf{H}_i'$.

The parameter $\theta \in \{0.3, 0.5\}$ governs regressor-error correlation and cross-equation correlation. The amount of serial correlation is specified through $\underline{\lambda}$ and $\bar{\lambda}$. The three scenarios $(\underline{\lambda}, \bar{\lambda}) \in \{(0.1, 0.5), (0.5, 0.8), (0.8, 0.95)\}$ steadily increase the autocorrelation in the generated data.

Setting C (Cointegration Tests): We continue to construct innovations according to Setting B. Moreover, we fix $\begin{bmatrix} \boldsymbol{\eta}_t \\ \boldsymbol{\varepsilon}_t \end{bmatrix} \stackrel{i.i.d.}{\sim} \mathbf{N}(\mathbf{0}, \boldsymbol{\Sigma}(\theta))$ with $\theta = 0.3$, and we construct the matrices \mathbf{A}_2 and \mathbf{A}_3 using $(\underline{\lambda}, \bar{\lambda}) = (0.1, 0.5)$. The eigenvalues of \mathbf{A}_1 are varied to explore both size and power properties. We always estimate a *quadratic* seemingly unrelated CPR.

Size DGP. Take the eigenvalues of \mathbf{A}_1 as before. That is, take $\lambda_{11}, \dots, \lambda_{1n} \stackrel{i.i.d.}{\sim} \mathbf{U}[\underline{\lambda}, \bar{\lambda}]$ with $(\underline{\lambda}, \bar{\lambda}) \in \{(0.1, 0.5), (0.5, 0.8), (0.8, 0.95)\}$.

Power DGP1. We set $\lambda_{1j} = 1$ for $1 \leq j \leq J_1$ and generate $\lambda_{1j} \stackrel{i.i.d.}{\sim} \mathbf{U}[0.1, 0.5]$ for $J_1 + 1 \leq j \leq n$. The integer $J_1 \in \{1, 2, n\}$ represents the number of unit roots in $\{\mathbf{u}_t\}$.

Power DGP2. The eigenvalues of Λ_1 are sampled as $\lambda_{11}, \dots, \lambda_{1n} \stackrel{i.i.d.}{\sim} U[0.1, 0.5]$, and the first $J_2 \in \{1, 2, n\}$ series follow a cubic SUCPR specification:

$$y_{it} = \begin{cases} 1 + t + 5x_{it} - 0.3x_{it}^2 + 0.01x_{it}^3 + u_{it}, & 1 \leq i \leq J_2, \\ 1 + t + 5x_{it} - 0.3x_{it}^2 + u_{it}, & J_2 + 1 \leq i \leq n. \end{cases}$$

Power DGP3. We again take $\lambda_{11}, \dots, \lambda_{1n} \stackrel{i.i.d.}{\sim} U[0.1, 0.5]$ and construct

$$y_{it} = \begin{cases} \sum_{s=1}^t u_{is}, & 1 \leq i \leq J_3, \\ 1 + t + 5x_{it} - 0.3x_{it}^2 + u_{it}, & J_3 + 1 \leq i \leq n, \end{cases}$$

where $J_3 \in \{1, 2, n\}$ represents for the number of equations that specify a spurious relation.

Overall, the Power DGPs 1-3 consider: missing $I(1)$ regressors, omitted higher order powers of the $I(1)$ regressor x_{it} , and spurious regressions, respectively.

3.4.2 Discussion of the Simulation Results

Tables 3.1 and 3.2 report the empirical mean squared error (MSE) for both feasible and infeasible estimators. As results are qualitatively similar across equations, we only report on the estimators for $\beta_{1,4}$ (the coefficient in front of x_{1t}^2). The column with FGLS contains the numerical value of the MSE and the MSEs of all other estimators are expressed relative to this benchmark. Values above 1 indicate a better performance of FM-GLS.

The FM-GLS estimator generally has the lowest MSE among all feasible estimators. These efficiency gains are small at low levels of endogeneity and serial correlation, but become sizeable at higher levels. Moreover, the Monte Carlo outcomes for the infeasible estimators indicate that these gains remain when the estimators are informed about the true endogeneity and serial

Table 3.1: Empirical MSE for the coefficient $\beta_{i,4}$ of x_h^2 with $i = 1$ under error Setting A. The column labeled FGLS contains the numerical value of the MSE of feasible FM-GLS. Other MSEs are expressed relative to this benchmark. Values above 1 indicate a better performance of feasible FM-GLS.

ρ	$n = 3$						$n = 5$					
	SOLS	SUR	FGLS	infSOLS	infSUR	infFGLS	SOLS	SUR	FGLS	infSOLS	infSUR	infFGLS
$T = 100$												
0	0.999	1.048	3.56E-05	0.937	0.937	0.937	0.988	1.047	3.81E-05	0.894	0.894	0.894
0.3	1.170	1.077	5.50E-05	1.098	0.959	0.907	1.186	1.064	5.43E-05	1.090	0.899	0.851
0.6	2.166	1.474	7.09E-05	2.017	1.180	0.864	2.260	1.426	6.85E-05	2.025	1.046	0.785
0.8	5.247	2.607	7.51E-05	6.547	2.474	0.878	5.720	2.589	7.40E-05	6.591	1.984	0.676
$T = 200$												
0	1.012	1.042	4.09E-06	0.961	0.961	0.961	1.019	1.069	4.25E-06	0.939	0.939	0.939
0.3	1.146	1.043	6.64E-06	1.101	0.960	0.942	1.216	1.069	6.62E-06	1.124	0.937	0.907
0.6	2.045	1.295	1.01E-05	1.920	1.101	0.906	2.222	1.345	9.37E-06	2.032	1.056	0.851
0.8	5.206	2.434	1.29E-05	5.502	1.971	0.916	6.197	2.701	1.11E-05	6.106	1.778	0.813
$T = 500$												
0	1.013	1.028	2.41E-07	0.988	0.988	0.988	1.016	1.039	2.55E-07	0.973	0.973	0.973
0.3	1.195	1.054	4.03E-07	1.162	0.998	0.975	1.224	1.052	4.02E-07	1.176	0.978	0.961
0.6	1.877	1.154	7.45E-07	1.789	1.054	0.952	2.142	1.217	6.49E-07	1.993	1.023	0.914
0.8	4.158	1.771	1.26E-06	4.011	1.456	0.935	5.341	2.031	1.03E-06	5.038	1.394	0.880

correlation properties. It is thus the GLS weighting of the data that improves estimation accuracy.

There is one particular instance in Table 3.2 in which the performance of the FM-GLS estimator has a high MSE, namely the case of high persistency $(\underline{\lambda}, \bar{\lambda}) = (0.8, 0.95)$, high endogeneity $\theta = 0.5$, and small sample size $T = 100$. This is caused by an inaccurate BIAM estimator resulting from the combination of a small sample size, high endogeneity, and high persistency. The problem disappears when T increases.

The subsequent set of simulations evolves around hypothesis testing, see Table 3.3 and Figures 3.1-3.3. The errors are simulated using Setting A and we use the following Wald-type test statistics: the Wald-SOLS and Wald-SUR tests as developed in Proposition 2 in Wagner et al. (2020), and the Wald-FGLS test from Theorem 3.4. We consider: (i) the single equation test $H_0 : \beta_{1,4} = -0.3$ against the two-sided alternative $H_1 : \beta_{1,4} \neq -0.3$, and (ii) the joint test $H_0 : \beta_{1,4} = \beta_{2,4} = \dots = \beta_{n,4} = -0.3$ against the alternative which rejects when at least one coefficient is unequal to -0.3 .

Table 3.2: Empirical MSE for the coefficient $\beta_{i,4}$ of x_{it}^2 with $i = 1$. under error Setting B. For further explanations see caption for Table 3.1.

$(\underline{\lambda}, \bar{\lambda})$	T	$n = 3$						$n = 5$					
		SOLS	SUR	FGLS	infSOLS	infSUR	infFGLS	SOLS	SUR	FGLS	infSOLS	infSUR	infFGLS
Panel A: Low endogeneity $\theta = 0.3$													
(0.1, 0.5)	100	1.290	1.270	9.46E-05	1.270	1.131	0.865	1.256	1.275	9.51E-05	1.215	1.052	0.822
	200	1.216	1.163	1.27E-05	1.191	1.077	0.916	1.220	1.164	1.22E-05	1.201	1.038	0.896
	500	1.137	1.088	8.58E-07	1.133	1.033	0.961	1.159	1.106	7.65E-07	1.151	0.997	0.936
(0.5, 0.8)	100	1.790	1.825	3.99E-05	1.792	1.522	0.638	1.812	1.896	3.81E-05	1.802	1.431	0.594
	200	1.677	1.681	4.91E-06	1.597	1.408	0.763	1.682	1.645	4.73E-06	1.581	1.271	0.733
	500	1.467	1.404	3.34E-07	1.392	1.248	0.897	1.522	1.415	3.01E-07	1.432	1.185	0.864
(0.8, 0.95)	100	0.123	0.136	1.48E-04	0.212	0.145	0.022	0.001	0.001	2.67E-02	0.002	0.001	0.000
	200	2.257	2.322	8.59E-07	2.537	1.986	0.517	1.622	1.577	1.04E-06	2.218	1.343	0.366
	500	2.098	2.050	5.47E-08	2.021	1.686	0.703	2.031	2.014	4.90E-08	1.964	1.394	0.632
Panel B: High endogeneity $\theta = 0.5$													
(0.1, 0.5)	100	1.327	1.269	7.48E-05	1.579	1.230	0.899	1.335	1.299	7.26E-05	1.825	1.211	0.817
	200	1.232	1.161	9.91E-06	1.394	1.131	0.949	1.269	1.225	9.26E-06	1.653	1.166	0.886
	500	1.179	1.091	6.66E-07	1.297	1.079	0.980	1.162	1.093	5.99E-07	1.464	1.071	0.942
(0.5, 0.8)	100	1.880	1.892	3.08E-05	2.732	1.861	0.661	1.850	1.833	3.05E-05	3.170	1.718	0.575
	200	1.793	1.727	3.72E-06	2.132	1.610	0.798	1.713	1.639	3.76E-06	2.143	1.381	0.697
	500	1.480	1.370	2.62E-07	1.678	1.299	0.941	1.577	1.461	2.44E-07	1.739	1.256	0.863
(0.8, 0.95)	100	0.030	0.031	5.14E-04	0.172	0.077	0.005	0.000	0.000	3.27E+03	0.000	0.000	0.000
	200	2.381	2.514	6.85E-07	5.002	2.825	0.541	1.880	1.771	7.83E-07	5.026	2.030	0.389
	500	2.186	2.090	4.23E-08	2.756	1.850	0.712	2.184	2.044	4.03E-08	2.786	1.560	0.619

For the size property, Table 3.3 shows the Wald tests are typically oversized but the three tests react differently to changes in ρ . Increases in ρ result in an increasing size for the SOLS and SUR version of the Wald test, whereas increases in ρ lead to size decreases for the GLS type of Wald test. For all three Wald-type of tests, the size of the tests improves with sample size T . Overall, the GLS test provides better size control. In Figure 3.1, we vary the serial correlation parameter ρ_1 and the endogeneity parameter ρ_2 separately. Overall, variation in ρ_1 has a larger influence on size with the SUR test being most sensitive and the GLS test being least sensitive.

Now we compare the finite-sample power property. The ordering in terms of size-corrected power is the same throughout Figures 3.2-3.3. That is, size-corrected power is lowest for Wald-SOLS, increases for Wald-SUR, and is highest for the Wald-FGLS test.

Finally, we the KPSS-type of cointegration tests. The results can be found in Table 3.4. The general conclusions are as follows. The empirical sizes of the K^{SOLS} and K^{SUR} tests are similar. We observe very conservative results for

Table 3.3: Empirical size (%) of the single-equation Wald tests for $H_0 : \beta_{1,4} = -0.3$ and the joint Wald tests for $H_0 : \beta_{1,4} = \beta_{2,4} = \dots = \beta_{n,4} = -0.3$, where $\beta_{i,4}$ denote the coefficients in front of x_{it}^2 .

	$n = 3$			$n = 5$		
ρ	Wald-SOLS	Wald-SUR	Wald-FGLS	Wald-SOLS	Wald-SUR	Wald-FGLS
Panel A: Single-equation test						
$T = 100$						
0	11.75	13.44	9.89	14.30	17.84	13.44
0.3	13.25	15.09	9.92	15.32	19.59	13.55
0.6	16.13	19.03	7.54	18.80	27.65	11.86
0.8	20.56	26.60	4.70	26.72	43.11	10.13
$T = 200$						
0	9.02	10.00	7.48	9.90	12.00	8.47
0.3	9.96	10.93	7.24	11.32	13.62	8.81
0.6	12.68	14.10	5.93	14.62	19.40	7.71
0.8	15.72	19.50	2.95	19.58	30.95	4.90
$T = 500$						
0	7.02	7.44	5.89	7.32	8.47	6.19
0.3	8.07	8.41	5.76	8.54	9.50	6.04
0.6	9.41	9.68	4.78	10.34	12.54	5.94
0.8	10.92	12.48	3.09	13.47	18.96	3.73
Panel B: Joint test						
$T = 100$						
0	17.85	21.68	14.60	29.13	39.40	26.65
0.3	20.67	24.62	14.44	32.60	44.91	28.14
0.6	27.13	33.59	11.40	45.65	65.44	27.72
0.8	36.99	48.49	8.30	63.29	86.38	26.57
$T = 200$						
0	11.66	13.74	9.00	17.49	22.90	13.26
0.3	14.11	16.25	8.77	21.35	28.23	14.55
0.6	19.08	22.71	7.10	30.58	44.07	12.88
0.8	26.21	33.92	4.19	44.82	69.44	10.60
$T = 500$						
0	8.70	9.64	6.60	10.83	13.38	7.91
0.3	9.84	10.74	6.23	13.29	16.19	8.19
0.6	12.79	14.09	5.13	18.69	25.15	7.67
0.8	16.02	19.56	3.35	26.86	41.72	5.28

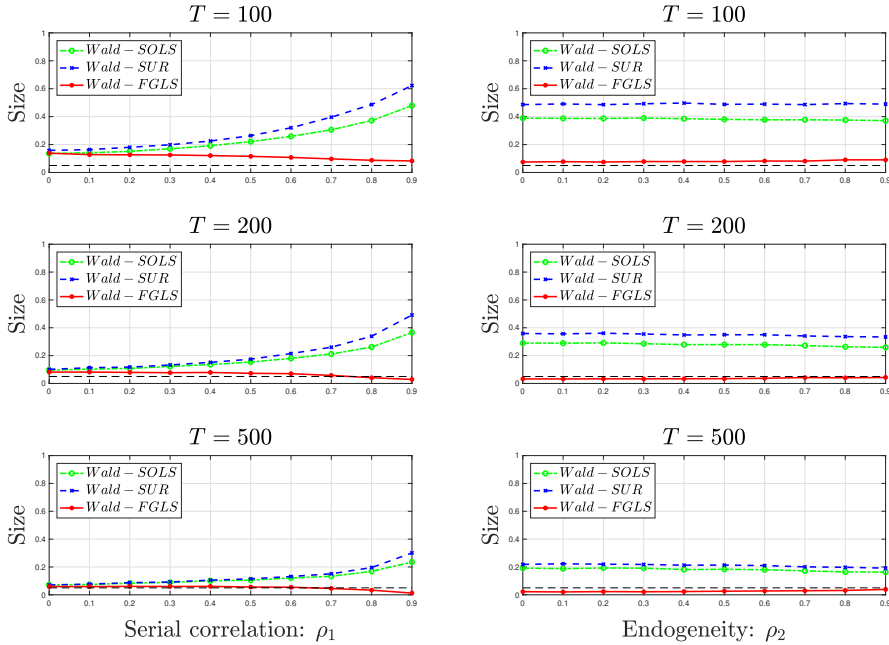
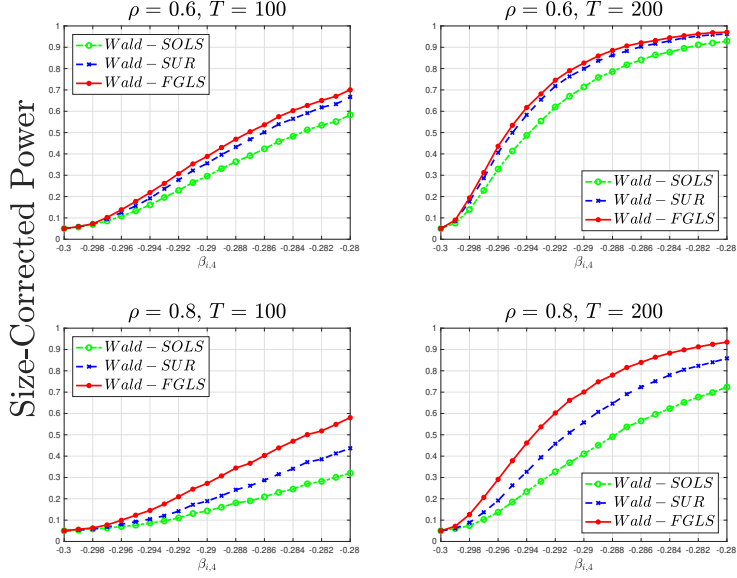


Figure 3.1: Empirical size of the joint Wald tests $H_0 : \beta_{1,4} = \beta_{2,4} = \dots = \beta_{n,4} = -0.3$, where $\beta_{i,4}$ denote the coefficients in front of x_{it}^2 . In the left part of the figure, we vary the serial correlation parameter ρ_1 from 0 to 0.9 while keeping $\rho_2 = \rho_3 = \rho_4 = 0.8$ fixed. Similarly, the endogeneity parameter ρ_2 is varied from 0 to 0.9 and $\rho_1 = \rho_3 = \rho_4 = 0.8$ remain unchanged. The cross-sectional dimension is $n = 3$ for both parts.

Single-equation tests, $n = 3$



Single-equation tests, $n = 5$

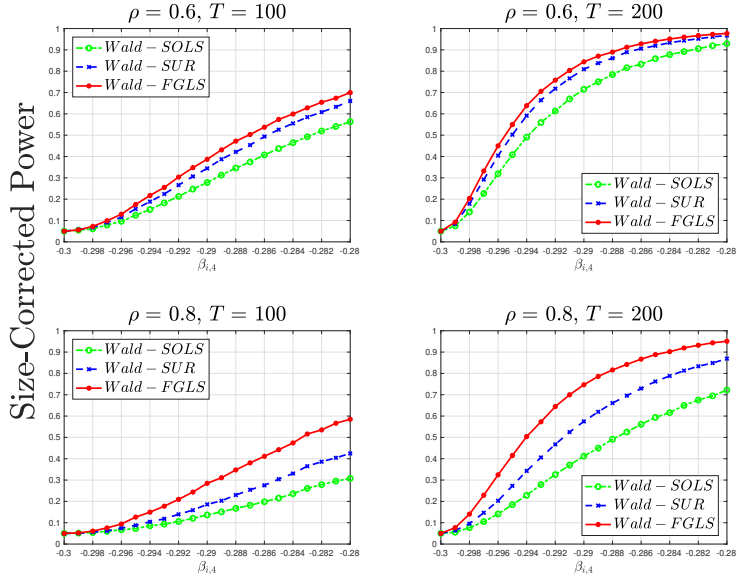
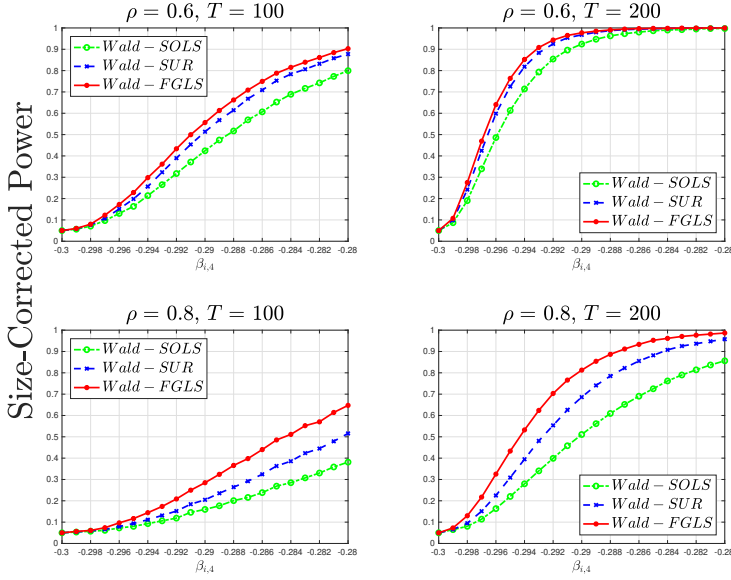


Figure 3.2: Empirical size-corrected power of the *single-equation* Wald tests $H_0 : \beta_{1,4} = -0.3$ where $\beta_{1,4}$ is the coefficient in front of x_{1t}^2 . We consider $n \in \{3, 5\}$, $T \in \{100, 200\}$, and $\rho_1 = \rho_2 = \rho_3 = \rho_4 = \rho \in \{0.6, 0.8\}$.

Joint tests, $n = 3$



Joint tests, $n = 5$

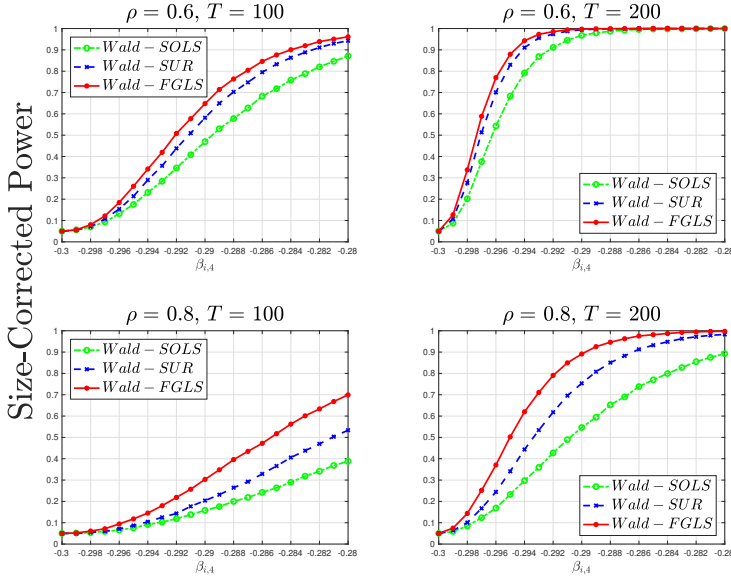


Figure 3.3: Empirical size-corrected power of the *joint* Wald tests $H_0 : \beta_{1,4} = \beta_{2,4} = \dots = \beta_{n,4} = -0.3$ where $\beta_{i,4}$ are the coefficients in front of x_{it}^2 . We consider $n \in \{3, 5\}$, $T \in \{100, 200\}$, and $\rho_1 = \rho_2 = \rho_3 = \rho_4 = \rho \in \{0.6, 0.8\}$.

Table 3.4: Empirical size (%) and power (%) of Bonferroni-type (multivariate) sub-sampling KPSS tests. The integer J_1 in Power DGP1 indicates the number of unit roots contained in errors $\{u_t\}$, J_2 is related to the number of equations that exclude cubic power terms x_{it}^3 , J_3 specifies the number of spurious relations.

		$n = 3$				$n = 5$		
		T	K^{SOLS}	K^{SUR}	K^{BIAM}	K^{SOLS}	K^{SUR}	K^{BIAM}
Panel A: Size								
$(\underline{\lambda}, \bar{\lambda})$, serial correlation	(0.1, 0.5)	100	0.37	0.34	0.47	0.33	0.25	0.26
		200	0.42	0.37	0.53	0.42	0.28	0.18
		500	0.71	0.60	1.24	0.64	0.58	0.80
	(0.5, 0.8)	100	8.10	7.28	1.86	22.35	19.65	1.29
		200	3.54	3.22	2.91	8.70	7.57	1.09
		500	1.92	1.69	4.54	4.02	3.40	4.78
	(0.8, 0.95)	100	50.63	48.67	16.17	90.99	89.69	18.41
		200	40.48	38.56	14.26	83.71	81.78	12.16
		500	21.87	20.36	16.25	60.09	56.87	15.49
Panel B: Power DGP1								
J_1 , # {unit roots}	1	100	44.03	39.92	21.14	57.65	52.05	14.00
		200	57.97	50.90	33.85	74.09	66.64	27.24
		500	67.23	57.37	56.63	88.41	80.23	49.93
	2	100	66.64	63.83	38.07	84.77	80.36	29.12
		200	77.77	73.00	52.88	94.38	91.16	49.87
		500	84.19	78.83	78.40	98.57	96.89	72.02
	n	100	78.89	77.41	54.26	99.39	99.24	64.89
		200	88.40	86.53	70.07	99.97	99.95	88.75
		500	91.28	90.19	89.55	100.00	100.00	96.53
Panel C: Power DGP2								
J_2 , # {misspecified equations}	1	100	10.92	10.51	3.47	18.67	18.40	1.91
		200	26.41	25.58	7.72	41.90	41.41	4.95
		500	55.72	55.28	24.10	77.76	77.22	18.36
	2	100	17.83	16.97	6.15	30.77	29.77	3.70
		200	40.29	39.04	15.57	62.01	61.15	10.51
		500	69.64	68.50	38.48	91.66	91.27	30.87
	n	100	23.16	22.53	9.88	53.66	51.21	10.45
		200	47.25	45.60	22.83	83.78	82.73	29.05
		500	75.60	73.98	56.79	98.23	98.15	69.35
Panel D: Power DGP3								
J_3 , # {spurious relations}	1	100	77.03	77.06	28.34	94.26	94.12	21.04
		200	89.55	89.14	35.64	98.97	98.84	31.79
		500	97.26	97.05	55.61	99.91	99.91	48.48
	2	100	87.40	86.75	52.85	98.95	98.81	43.05
		200	94.95	94.33	57.95	99.89	99.89	58.15
		500	98.69	98.49	77.59	100.00	100.00	69.59
	n	100	91.55	90.60	71.95	99.97	99.92	87.87
		200	96.71	96.33	74.85	99.99	100.00	93.86
		500	98.04	97.62	91.66	100.00	100.00	98.02

Note: To decrease the computational burden we reduced the number of Monte Carlo replications to 10^4 .

low serial correlation, decent size for medium serial correlation, and strongly oversized tests at high levels of serial correlation. These findings are completely in line with the simulation results that are reported in table 3 of Choi and Saikkonen (2010). The same behaviour is observed for the K^{BIAM} test but the deviations from the 5% level are less extreme. The power of the K^{FOLS} , K^{SUR} , and K^{BIAM} tests behaves as expected: (1) power always increases with sample size, and (2) power increases when more unit roots, more misspecified equations, or more spurious relationships are incorporated in the DGP. The K^{BIAM} test has the lowest power among the three tests. This is caused by the fact that the filter can nearly difference the data and hence make it appear more stationary.

3.5 Empirical Application

The Environmental Kuznets Curve (EKC) conjectures an inverted U-shaped relation between environmental degradation and income per capita. That is, there is an initial decline in environmental quality with increasing economic activity, but beyond a certain turning point (caused by e.g. industrial transformation and increasing environmental awareness), economic growth goes hand in hand with environmental improvement. A more detailed description and historical overview of the EKC can be found in Stern (2004) and Stern (2017), respectively. The implications of further economic growth on pollution, e.g. the emission of greenhouse-gases, are also key in understanding the future of global warming (Nordhaus (2013)).

We build upon and compare with Wagner et al. (2020). That is, we look at carbon dioxide (CO_2) emissions and GDP as proxies for environmental pollution and economic development (both per capita and in logarithms), respectively. The data is collected from the Maddison Project Database (MPD) and the homepage of the Carbon Dioxide Information Analysis Center (CDIAC).⁸ As

⁸The Maddison Project Database, Bolt et al. (2018), contains the data on population size and real GDP. The data on CO_2 originates from Boden et al. (2017). We follow the official guidelines and multiply by 3.667 and 10^6 to convert the reported fossil-fuel emissions into

Table 3.5: Outcomes for the joint tests of cointegration in the empirical study. At a significance level of 5%, the null hypothesis of cointegration is rejected when ‘rejection rule’ is less than 5%.

	K^{SOLS}	K^{SUR}	K^{BIAM}
Statistic	16.54	12.66	8.19
Rejection Rule (in %)	0.00	0.03	2.73
Block Size	22	20	22
#{blocks}	6	7	6

in Wagner et al. (2020), we consider Austria (AT), Belgium (BE), Finland (FI), the Netherlands (NL), Switzerland (CH) and the United Kingdom (UK). Our yearly data spans the period from 1870 to 2014. We refer to the latter paper for a discussion of the stationarity properties of all series as well as the motivation for this particular set of countries. Overall, the dataset consist of $n = 6$ countries with $T = 145$ time series observations each. Such a panel with small n and large T is ideally suited for our FMGLS approach since the multivariate banded inverse autocovariance matrix remains computable.

We estimate the following quadratic model:

$$e_{it} = \beta_{i,1} + \beta_{i,2}t + \beta_{i,3}g_{it} + \beta_{i,4}g_{it}^2 + u_{it}, \quad (3.33)$$

where e_{it} are CO₂ emissions, and g_{it} are GDP. As the first step in our analysis we employ the multivariate stationarity tests of Section 3.3.4 to check this model specification (Table 3.5 above). All three tests reject the null of cointegration at a 5% level signalling inappropriateness of the quadratic formulation. Figure 3.4 shows the residuals on which these tests are based. What stands out in these graphs is the erratic behaviour of the series around the two world wars. Based on this fact, and to be able to compare to Wagner et al. (2020), we will continue the analysis using model (3.33) and the given collection of countries. Before doing so, it will be worthwhile to discuss the time series properties of these residuals.

total carbon dioxide emissions.

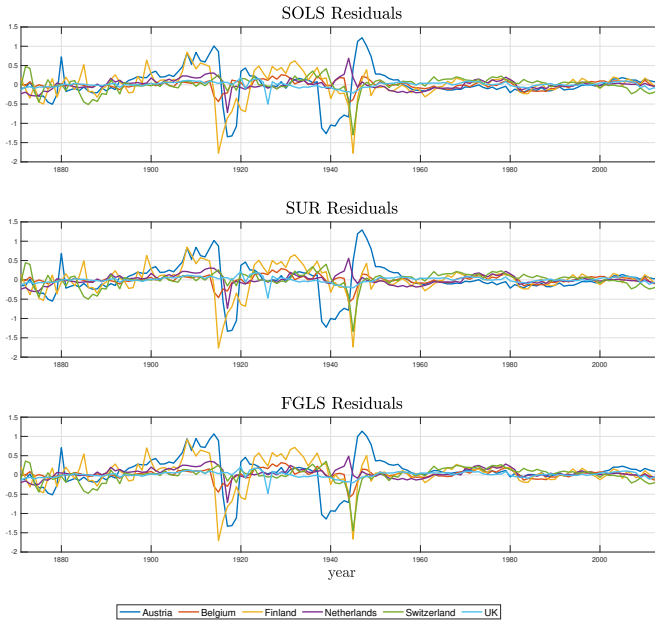


Figure 3.4: The plots of the residuals $\hat{u}_{t,SOLS} = y_t - Z_t \hat{\beta}_{SOLS}^*$ (top), $\hat{u}_{t,SUR} = y_t - Z_t \hat{\beta}_{SUR}^*$ (middle), and $\hat{u}_{t,FGLS} = y_t - Z_t \hat{\beta}_{FGLS}^*$ (bottom) for the empirical study.

We consider the series $\{\hat{u}_{t,FGLS}\}$ in the remainder of this section but the other residuals series will provide qualitatively similar outcomes. When fitting the $VAR(p)$ models with $1 \leq p \leq 8$ to these residuals, the BIC information criterion selects a lag order of $p = 1$. The absolute eigenvalues of the estimated coefficient matrix are $(0.55, 0.55, 0.51, 0.31, 0.31, 0.11)$, and the estimate for the error correlation matrix is

$$\begin{array}{c}
 \begin{array}{ccccc}
 & AT & BE & FI & NL & CH & UK \\
 AT & 1 & 0.16 & 0.10 & 0.16 & 0.18 & 0.22 \\
 BE & \bullet & 1 & 0.51 & 0.09 & 0.23 & 0.27 \\
 FI & \bullet & \bullet & 1 & 0.13 & 0.26 & 0.22 \\
 NL & \bullet & \bullet & \bullet & 1 & 0.22 & 0.10 \\
 CH & \bullet & \bullet & \bullet & \bullet & 1 & 0.18 \\
 UK & \bullet & \bullet & \bullet & \bullet & \bullet & 1
 \end{array}
 \end{array}
 .$$

There is thus serial and cross-sectional correlation to be exploited by the FM-GLS estimator.

The FM-SOLS, FM-SUR and FM-GLS estimation results of Model (3.33) are reported in Table 3.6. An inspection of the coefficient estimates and their confidence intervals reveals that: (1) $\beta_{i,3}$ is positive for each country, (2) $\beta_{i,4}$ is negative for each country, and (3) all coefficients are significant at the 5% level. All these three facts are in line with the EKC hypothesis.⁹ Accordingly, there exists a turning point after which further per capita economic growth reduces per capita carbon dioxide emissions. The numerical values for the turning points are heterogeneous between countries.

The widths of the confidence intervals for $\beta_{i,3}$ and $\beta_{i,4}$ display a similar pattern. From shortest to longest, the ordering is always FM-SUR, FM-SOLS, and FM-GLS and we also see how widths vary substantially between methods. To uncover the origin of these findings we conduct one final simulation study with a parameter specification that closely mimics the properties of the dataset.¹⁰ The average empirical coverage probabilities of asymptotic 95% confidence intervals are 78.0%, 66.5% and 89.0% for FM-SOLS, FM-SUR, and FM-GLS, respectively. In other words, the calculated confidence intervals are generally too short. By reverse engineering it turns out that the confidence intervals should be scaled by factors of 1.67, 2.15 and 1.24 to bring them back to the desired nominal level. Overall, the applied researcher should be careful when using the confidence intervals as indications for parameter uncertainty.

⁹This is non-surprising because Wagner et al. (2020) have selected the current set of countries because they display the EKC behaviour. Also, our estimation results are slightly different from those in Wagner et al. (2020) due to the additional data for 2014, possible data updates, and/or differences in the bandwidth selection of the long-run covariance matrices.

¹⁰The details of this simulation DGP are provided in Appendix F. A visualisation of the data and the model fit are also provided there.

Table 3.6: Estimates for $\beta_{i,3}$ and $\beta_{i,4}$ from the quadratic EKC model in (3.33). The numbers between parentheses are 95% asymptotic confidence intervals. Turning points are computed as $\exp(-\widehat{\beta}_{i,3}/2\widehat{\beta}_{i,4})$.

		x_t	x_t^2	Turning point
Austria	SOLS	9.173 (6.797,11.548)	-0.411 (-0.535,-0.288)	68,900
	SUR	8.494 (6.764,10.223)	-0.370 (-0.464,-0.276)	76,211
	FGLS	6.553 (2.138,10.967)	-0.276 (-0.513,-0.040)	708,712
Belgium	SOLS	12.927 (11.795,15.059)	-0.645 (-0.702,-0.589)	22,420
	SUR	9.973 (9.158,10.787)	-0.503 (-0.545,-0.461)	20,195
	FGLS	8.762 (7.236,10,287)	-0.443 (-0.521,-0.365)	19,795
Finland	SOLS	15.676 (14.162,17.289)	-0.716 (-0.788,-0.643)	56,967
	SUR	15.136 (14.030,16.242)	-0.684 (-0.742,-0.627)	63,400
	FGLS	14.392 (12.075,16.708)	-0.646 (-0.769,-0.523)	68,775
Netherlands	SOLS	11.382 (10.140,12.624)	-0.540 (-0.602,-0.477)	38,120
	SUR	10.063 (9.183,10.944)	-0.475 (-0.522,-0.429)	39,637
	FGLS	9.102 (7.606,10.597)	-0.430 (-0.506,-0.353)	39,908
Switzerland	SOLS	7.070 (5.516,8.624)	-0.232 (-0.310,-0.153)	4.287×10^6
	SUR	6.962 (5.481,8.443)	-0.232 (-0.308,-0.156)	3.316×10^6
	FGLS	7.052 (5.173,8.932)	-0.254 (-0.350,-0.158)	1.074×10^6
United Kingdom	SOLS	8.450 (6.976,10.020)	-0.429 (-0.502,-0.355)	20,242
	SUR	9.523 (8.486,10.560)	-0.475 (-0.527,-0.423)	22,596
	FGLS	9.056 (7.244,10.868)	-0.453 (-0.542,-0.364)	21,887

3.6 Conclusion

We proposed a framework to conduct inference on cointegrating polynomial regressions. Parameters are obtained using a Fully Modified GLS estimator and we studied a cointegration test that is based on filtered residuals. Monte Carlo simulations revealed the advantages and disadvantages of these methods. The empirical researcher should realize that all estimation approaches have a tendency to underestimate parameter uncertainty and thus provide confidence intervals that are too small. The FM-GLS estimator suffers the least from this problem. Several interesting questions are left for future research. From a theoretical viewpoint, it is interesting to study the behaviour of the modified Cholesky decomposition (and BIAM) when the series under consideration is nonstationary. This would give insights into the behaviour of: (1) the FM-GLS estimator while estimating spurious regressions, and (2) the power properties of the cointegration tests. From a practical viewpoint, there seems a need to obtain more accurate standard errors of the parameter estimators.

Appendix 3.A Auxiliary Results

In this section, we show some preliminary lemmas that are used to establish the main results in this chapter.

Lemma 3.1

If $\{\mathbf{u}_t\}$ satisfies Assumption 3.1, then for any $m \geq 1$, there exists a constant $C > 0$ such that

$$\sum_{j=1}^m \|\mathbf{A}_j(m) - \mathbf{A}_j\|_{\mathcal{F}}^2 \leq C \sum_{j=m+1}^{\infty} \|\mathbf{A}_j\|_{\mathcal{F}}^2. \quad (3.34)$$

Proof In view of page 257 of Hannan and Deistler (2012), the summability condition of Assumption 3.1 implies that the spectral density matrix is bounded and bounded away from zero. The boundedness condition in Cheng and Pourahmadi (1993) is thus satisfied and (3.34) follows from their Theorem 2.2. ■

Lemma 3.2 (Implications of the First Moment Bound Theorem)

Let Assumption 3.1 hold, and define

$$\boldsymbol{\eta}_{t+1,\ell} = \mathbf{u}_{t+1} - \mathbf{A}(\ell)\mathbf{u}_t(\ell), \quad \text{where } \mathbf{u}_t(\ell) = [\mathbf{u}'_t, \mathbf{u}'_{t-1}, \dots, \mathbf{u}'_{t-\ell+1}]'. \quad (3.35)$$

The following three inequalities are true:

- (a) $\mathbb{E} \left\| \frac{1}{T-q} \sum_{t=q}^{T-1} \mathbf{u}_t(q) \mathbf{u}_t(q)' - \mathbb{E}(\mathbf{u}_t(q) \mathbf{u}_t(q)') \right\|^2 \leq C \frac{q^2}{T-q};$
- (b) $\mathbb{E} \left\| \frac{1}{T-\ell} \sum_{t=\ell}^{T-1} (\boldsymbol{\eta}_{t+1,\ell} - \boldsymbol{\eta}_{t+1}) \mathbf{u}_t(\ell)' \right\|^r \leq C \left(\frac{\ell}{T-\ell} \right)^{r/2} \left(\sum_{j=\ell+1}^{\infty} \|\mathbf{A}_j\|_{\mathcal{F}}^2 \right)^{r/2},$ for some $r \geq 2$ and any $1 \leq \ell \leq q$;
- (c) $\mathbb{E} \left\| \frac{1}{T-\ell} \sum_{t=\ell}^{T-1} (\boldsymbol{\eta}_{t+1,\ell} \boldsymbol{\eta}'_{t+1,\ell}) - \mathbb{E}(\boldsymbol{\eta}_{t+1,\ell} \boldsymbol{\eta}'_{t+1,\ell}) \right\|^r \leq C(T-\ell)^{-r/2},$ for some $r \geq 2$ and any $1 \leq \ell \leq q$.

Proof (a) Since $\|\cdot\|^2 \leq \|\cdot\|_{\mathcal{F}}^2$, we obtain

$$\begin{aligned} & \mathbb{E} \left\| \frac{1}{T-q} \sum_{t=q}^{T-1} \mathbf{u}_t(q) \mathbf{u}_t(q)' - \mathbb{E}(\mathbf{u}_t(q) \mathbf{u}_t(q)') \right\|^2 \\ & \leq \frac{1}{(T-q)^2} \sum_{k,\ell=1}^q \sum_{i,j=1}^n \mathbb{E} \left[\sum_{t=q}^{T-1} \left[u_{i,t-k+1} u_{j,t-\ell+1} - \mathbb{E}(u_{i,t-k+1} u_{j,t-\ell+1}) \right] \right]^2, \end{aligned} \quad (3.36)$$

where $u_{i,t}$ denotes the i^{th} element \mathbf{u}_t . As remarked in the main text, the lag polynomial $\mathcal{A}(L)$ is invertible by Assumption 3.1. Recall $C(L) = [\mathcal{A}(L)]^{-1} = \sum_{j=0}^{\infty} C_j L^j$ with $C_0 = \mathbf{I}_n$, and $\sum_{j=0}^{\infty} j \|C_j\|_{\mathcal{F}} < \infty$. We observe that $u_{i,t} = \sum_{j=0}^{\infty} \text{row}_i(C_j) \boldsymbol{\eta}_{t-j}$. By Proposition 10.2(b) of Hamilton (1994), absolute summability of the coefficient matrices $\{C_j\}_{j=0}^{\infty}$ implies $\sum_{s=0}^{\infty} |\gamma_{u,k}(s)| < \infty$ where $\gamma_{u,k}(s) = \mathbb{E}(u_{k,t} u_{k,t-s})$. The conditions for the First Moment Bound Theorem (FMBT) in Findley and Wei (1993) are thus satisfied. Choosing $q(t, s) = 1$ if $t = s \geq q$ (the banding parameter) and $q(t, s) = 0$ otherwise, by the FMBT,

$$\begin{aligned} & \mathbb{E} \left[\sum_{t=q}^{T-1} \left[u_{i,t-k+1} u_{j,t-\ell+1} - \mathbb{E}(u_{i,t-k+1} u_{j,t-\ell+1}) \right] \right]^2 \\ & = \mathbb{E} \left[\sum_{t,s=1}^{T-1} q(t, s) \left[u_{i,t-k+1} u_{j,s-\ell+1} - \mathbb{E}(u_{i,t-k+1} u_{j,s-\ell+1}) \right] \right]^2 \\ & \leq C \sum_{t,s,l,w=1}^{T-1} q(t, s) q(l, w) \gamma_{u,i}(t-l) \gamma_{u,j}(s-w) \\ & = C \sum_{t,l=q}^{T-1} \gamma_{u,i}(t-l) \gamma_{u,j}(t-l) \\ & \leq C \gamma_{u,j}(0) \sum_{t,l=q}^{T-1} |\gamma_{u,i}(t-l)| \leq C(T-q) \left[\gamma_{u,j}(0) \sum_{t=-\infty}^{\infty} |\gamma_{u,i}(t)| \right] \leq C(T-q). \end{aligned}$$

This bound holds for general k, ℓ, i and j , and **(a)** thereby follows from (3.36).

(b) For $1 \leq \ell \leq q$ and $r \geq 2$, we have

$$\mathbb{E} \left\| \frac{1}{T-\ell} \sum_{t=\ell}^{T-1} (\boldsymbol{\eta}_{t+1,\ell} - \boldsymbol{\eta}_{t+1}) \mathbf{u}_t(\ell)' \right\|^r \leq \ell^{r/2-1} (T-\ell)^{-r} \sum_{s=0}^{\ell-1} \mathbb{E} \left\| \sum_{t=\ell}^{T-1} (\boldsymbol{\eta}_{t+1,\ell} - \boldsymbol{\eta}_{t+1}) \mathbf{u}_{t-s}' \right\|_{\mathcal{F}}^r, \quad (3.37)$$

by $\|\cdot\|^r \leq (\|\cdot\|_{\mathcal{F}}^2)^{r/2}$ and the c_r -inequality. By assumption, $\boldsymbol{\eta}_{t+1}$ is uncorrelated with $[\mathbf{u}_{t-\ell+1}', \dots, \mathbf{u}_t']'$ implying that $\mathbb{E}[\sum_{t=\ell}^{T-1} (\boldsymbol{\eta}_{t+1,\ell} - \boldsymbol{\eta}_{t+1}) \mathbf{u}_{t-s}'] = \mathbf{O}$. The FMBT can

thus be applied directly without having to express the quadratic form in deviations from the mean. However, some rewriting is needed to obtain expressions in scalar random sequences. To this end, use $A_{j,kl}$ and $A_{j,kl}(\ell)$ to denote the $(k, l)^{th}$ element of \mathbf{A}_j and $\mathbf{A}_j(\ell)$, respectively. Setting $\mathbf{A}_j(\ell) = \mathbf{O}$ for $j > \ell$, we have $\boldsymbol{\eta}_{t+1,\ell} - \boldsymbol{\eta}_{t+1} = \sum_{j=1}^{\infty} [\mathbf{A}_j - \mathbf{A}_j(\ell)] \mathbf{u}_{t+1-j}$, and hence

$$\begin{aligned}
 & \mathbb{E} \left\| \sum_{t=\ell}^{T-1} (\boldsymbol{\eta}_{t+1,\ell} - \boldsymbol{\eta}_{t+1}) \mathbf{u}'_{t-s} \right\|_{\mathcal{F}}^r \\
 &= \mathbb{E} \left\{ \sum_{k,m=1}^n \left[\sum_{t=\ell}^{T-1} u_{m,t-s} \sum_{j=1}^{\infty} \sum_{l=1}^n (A_{j,kl} - A_{j,kl}(\ell)) u_{l,t+1-j} \right]^2 \right\}^{r/2} \\
 &\leq n^{r-2} \sum_{k,m=1}^n \mathbb{E} \left| \sum_{t=\ell}^{T-1} u_{m,t-s} \sum_{j=1}^{\infty} \sum_{l=1}^n (A_{j,kl} - A_{j,kl}(\ell)) u_{l,t+1-j} \right|^r \\
 &= n^{r-2} \sum_{k,m=1}^n \mathbb{E} \left| \sum_{t=\ell}^{T-1} u_{m,t-s} u_t^* \right|^r \tag{3.38}
 \end{aligned}$$

with $u_t^* = \sum_{j=1}^{\infty} \sum_{l=1}^n (A_{j,kl} - A_{j,kl}(\ell)) u_{l,t+1-j}$, where we suppress the dependence on the index k (also below) without confusion. To apply the FMBT, we define the autocovariances $\gamma_{u^*}(t-h) = \mathbb{E}(u_t^* u_h^*)$, the difference in lag polynomial coefficients $\mathbf{a}_l(\ell) = [A_{1,kl} - A_{1,kl}(\ell), A_{2,kl} - A_{2,kl}(\ell), \dots]'$ and $\boldsymbol{\Sigma}_{u_i, \infty} = [\gamma_{u_i, \infty}(i-j), 1 \leq i, j < \infty]$. By the Cauchy-Schwartz inequality, the c_r -inequality, and boundedness of the maximum eigenvalue of $\boldsymbol{\Sigma}_{u_i, \infty}$, we obtain

$$\begin{aligned}
 \gamma_{u^*}(t-h) &\leq \gamma_{u^*}(0) = \mathbb{E} \left[\sum_{l=1}^n \sum_{j=1}^{\infty} (A_{j,kl} - A_{j,kl}(\ell)) u_{l,t+1-j} \right]^2 \\
 &\leq n \sum_{l=1}^n [\mathbf{a}_{kl}(\ell)' \boldsymbol{\Sigma}_{u_l, \infty} \mathbf{a}_{kl}(\ell)] \\
 &\leq Cn \sum_{l=1}^n \|\mathbf{a}_{kl}(\ell)\|^2 \\
 &\leq Cn \sum_{j=1}^{\infty} \|\mathbf{A}_j - \mathbf{A}_j(\ell)\|_{\mathcal{F}}^2 \\
 &\leq Cn \sum_{j=\ell+1}^{\infty} \|\mathbf{A}_j\|_{\mathcal{F}}^2. \tag{3.39}
 \end{aligned}$$

Applying the FMBT, we have

$$\begin{aligned}
\left(\mathbb{E} \left[\sum_{t=\ell}^{T-1} |u_{m,t-s} u_t^*| \right]^r \right)^{1/r} &\leq C \left[\sum_{t,h=\ell}^{T-1} \gamma_{u,m}(t-h) \gamma_{u^*}(t-h) \right]^{1/2} \\
&\leq C \left[\gamma_{u^*}(0)(T-\ell) \sum_{t=-\infty}^{\infty} |\gamma_{u,m}(t)| \right]^{1/2} \\
&\leq C(T-\ell)^{1/2} \left(\sum_{j=\ell+1}^{\infty} \|\mathbf{A}_j\|_{\mathcal{F}}^2 \right)^{1/2}, \tag{3.40}
\end{aligned}$$

using (3.39) and the absolute summability of $\{\gamma_{u,m}(t)\}$. Combining (3.37), (3.38) and (3.40) leads to the desired inequality.

(c) The equality $\boldsymbol{\eta}_{t+1,\ell} = (\mathbf{I}_n - \sum_{j=1}^{\ell} \mathbf{A}_j(\ell)L^j)\mathbf{C}(L)\boldsymbol{\eta}_{t+1}$ shows that $\boldsymbol{\eta}_{t+1,\ell}$ has a linear process representation in terms of $\boldsymbol{\eta}_t$. Theorem 6.6.12 of Hannan and Deistler (2012) implies that $\sup_{1 \leq \ell < \infty} \sum_{j=0}^{\ell} \|\mathbf{A}_j(\ell)\|_{\mathcal{F}} < \infty$. By Propositions 10.2(b) and 10.3 of Hamilton (1994), both the coefficient matrices associated with $(\mathbf{I}_n - \sum_{j=1}^{\ell} \mathbf{A}_j(\ell)L^j)\mathbf{C}(L)$ and the autocovariances $\{\mathbb{E}(\eta_{k,t+1,\ell}\eta_{k,t+1-s,\ell})\}_{s=0}^{\infty}$ are absolutely summable, where $\eta_{k,t+1,\ell}$ is the k^{th} entry of $\boldsymbol{\eta}_{t+1,\ell}$. The proof is completed using the c_r -inequality and the FMBT, that is, for $r \geq 2$,

$$\begin{aligned}
&\mathbb{E} \left\| \frac{1}{T-\ell} \sum_{t=\ell}^{T-1} (\boldsymbol{\eta}_{t+1,\ell} \boldsymbol{\eta}_{t+1,\ell}' - \mathbb{E}(\boldsymbol{\eta}_{t+1,\ell} \boldsymbol{\eta}_{t+1,\ell}')) \right\|^r \\
&\leq n^{r-2} \frac{1}{(T-\ell)^r} \sum_{k,m=1}^n \mathbb{E} \left| \sum_{t=\ell}^{T-1} [\eta_{k,t+1,\ell} \eta_{m,t+1,\ell} - \mathbb{E}(\eta_{k,t+1,\ell} \eta_{m,t+1,\ell})] \right|^r \\
&\leq C n^{r-2} \frac{1}{(T-\ell)^r} \sum_{k,m=1}^n \left[\sum_{t,l=\ell}^{T-1} \mathbb{E}(\eta_{k,t+1,\ell} \eta_{k,l+1,\ell}) \mathbb{E}(\eta_{m,t+1,\ell} \eta_{m,l+1,\ell}) \right]^{r/2} \\
&\leq C \frac{1}{(T-\ell)^{r/2}}.
\end{aligned}$$

■

Lemma 3.3

If Assumptions 3.1-3.3 hold, then

$$\max_{1 \leq \ell \leq q} \|\widehat{\mathbf{A}}(\ell) - \mathbf{A}(\ell)\| = O_p\left(\frac{q}{\sqrt{T}}\right), \quad \max_{1 \leq \ell \leq q} \|\widehat{\mathbf{S}}(\ell) - \mathbf{S}(\ell)\| = O_p\left(\frac{q}{\sqrt{T}}\right).$$

Proof Recall the definition of $\eta_{t+1,\ell}$ and $u_t(\ell)$ in (3.35). Similarly, define

$$\widehat{\eta}_{t+1,\ell} = \widehat{u}_{t+1} - \widehat{A}(\ell)\widehat{u}_t(\ell), \quad \widetilde{\eta}_{t+1,\ell} = \widehat{u}_{t+1} - \widehat{A}(\ell)\widehat{u}_t(\ell), \quad (3.41)$$

and $\widehat{u}_t(\ell) = [\widehat{u}_t', \widehat{u}_{t-1}', \dots, \widehat{u}_{t-\ell+1}']'$. We first consider $\max_{1 \leq \ell \leq q} \|\widehat{A}(\ell) - A(\ell)\|$. Since $(\widehat{A}(\ell) - A(\ell))\widehat{u}_t(\ell) = \widehat{\eta}_{t+1,\ell} - \widetilde{\eta}_{t+1,\ell}$ and $\frac{1}{T-\ell} \sum_{t=\ell}^{T-1} \widetilde{\eta}_{t+1,\ell}\widehat{u}_t(\ell)' = \mathbf{O}$ (the first-order condition from (3.15)), we have

$$(\widehat{A}(\ell) - A(\ell)) \left(\frac{1}{T-\ell} \sum_{t=\ell}^{T-1} \widehat{u}_t(\ell)\widehat{u}_t(\ell)' \right) = \frac{1}{T-\ell} \sum_{t=\ell}^{T-1} \widehat{\eta}_{t+1,\ell}\widehat{u}_t(\ell)'. \quad (3.42)$$

If we can show that $\frac{1}{T-q} \sum_{t=q}^{T-1} \widehat{u}_t(q)\widehat{u}_t(q)'$ is asymptotically invertible, then $\frac{1}{T-\ell} \sum_{t=\ell}^{T-1} \widehat{u}_t(\ell)\widehat{u}_t(\ell)'$ must also be asymptotically invertible with probability 1.¹¹ By the triangular inequality, $\left\| \frac{1}{T-q} \sum_{t=q}^{T-1} \widehat{u}_t(q)\widehat{u}_t(q)' - \mathbb{E}(u_t(q)u_t(q)') \right\| \leq I_a + I_b$, where

$$I_a = \left\| \frac{1}{T-q} \sum_{t=q}^{T-1} u_t(q)u_t(q)' - \mathbb{E}(u_t(q)u_t(q)') \right\| = O_p \left(\frac{q}{\sqrt{T}} \right)$$

by Chebyshev's inequality and Lemma 3.2 (i), and

$$\begin{aligned} I_b &= \left\| \frac{1}{T-q} \sum_{t=q}^{T-1} \widehat{u}_t(q)\widehat{u}_t(q)' - \frac{1}{T-q} \sum_{t=q}^{T-1} u_t(q)u_t(q)' \right\| \\ &\leq \frac{1}{T-q} \sum_{t=q}^{T-1} \|\widehat{u}_t(q) - u_t(q)\|^2 + 2 \sqrt{\frac{1}{T-q} \sum_{t=q}^{T-1} \|\widehat{u}_t(q) - u_t(q)\|^2} \sqrt{\frac{1}{T-q} \sum_{t=q}^{T-1} \|u_t(q)\|^2}. \end{aligned}$$

We have $\frac{1}{T-q} \sum_{t=q}^{T-1} \|\widehat{u}_t(q) - u_t(q)\|^2 = \frac{1}{T-q} \sum_{t=q}^{T-1} \sum_{s=t-q+1}^t \|\widehat{u}_s - u_s\|^2 \leq \frac{q}{T-q} \|\widehat{u} - u\|^2 = \frac{q}{T} O_p(1)$ by Assumption 3.2. Because $\frac{1}{T-q} \sum_{t=q}^{T-1} \|u_t(q)\| = O_p(q)$ by Markov's inequality, we conclude $I_b = O_p(q/\sqrt{T})$. Overall, this gives

$$\left\| \frac{1}{T-q} \sum_{t=q}^{T-1} \widehat{u}_t(q)\widehat{u}_t(q)' - \mathbb{E}(u_t(q)u_t(q)') \right\| = O_p \left(\frac{q}{\sqrt{T}} \right) = o_p(1). \quad (3.43)$$

Now observe that $\mathbb{E}(u_t(q)u_t(q)')$ is a leading principal submatrix of Σ_u . As a result, $\frac{1}{T-q} \sum_{t=q}^{T-1} \widehat{u}_t(q)\widehat{u}_t(q)'$ is asymptotically invertible.

We subsequently bound the RHS of (3.42) as follows:

¹¹If the matrix Q is invertible, then each leading principle submatrix of Q is invertible as well.

$\max_{1 \leq \ell \leq q} \left\| \frac{1}{T-\ell} \sum_{t=\ell}^{T-1} \widehat{\boldsymbol{\eta}}_{t+1,\ell} \widehat{\mathbf{u}}_t(\ell)' \right\| \leq II_a + \dots + II_e$, where

$$\begin{aligned} II_a &= \max_{1 \leq \ell \leq q} \left\| \frac{1}{T-\ell} \sum_{t=\ell}^{T-1} \boldsymbol{\eta}_{t+1} \mathbf{u}_t(\ell)' \right\|, \\ II_b &= \max_{1 \leq \ell \leq q} \left\| \frac{1}{T-\ell} \sum_{t=\ell}^{T-1} (\boldsymbol{\eta}_{t+1,\ell} - \boldsymbol{\eta}_{t+1}) \mathbf{u}_t(\ell)' \right\|, \\ II_c &= \max_{1 \leq \ell \leq q} \left\| \frac{1}{T-\ell} \sum_{t=\ell}^{T-1} (\widehat{\boldsymbol{\eta}}_{t+1,\ell} - \boldsymbol{\eta}_{t+1,\ell}) \mathbf{u}_t(\ell)' \right\|, \\ II_d &= \max_{1 \leq \ell \leq q} \left\| \frac{1}{T-\ell} \sum_{t=\ell}^{T-1} \boldsymbol{\eta}_{t+1,\ell} (\widehat{\mathbf{u}}_t(\ell) - \mathbf{u}_t(\ell))' \right\|, \end{aligned}$$

$$\text{and } II_e = \max_{1 \leq \ell \leq q} \left\| \frac{1}{T-\ell} \sum_{t=\ell}^{T-1} (\widehat{\boldsymbol{\eta}}_{t+1,\ell} - \boldsymbol{\eta}_{t+1,\ell}) (\widehat{\mathbf{u}}_t(\ell) - \mathbf{u}_t(\ell))' \right\|.$$

We consider these terms separately starting from II_a . Using the properties of Frobenius norm,

$$\left\| \frac{1}{T-\ell} \sum_{t=\ell}^{T-1} \boldsymbol{\eta}_{t+1} \mathbf{u}_t(\ell)' \right\|^2 \leq \frac{1}{(T-\ell)^2} \sum_{s=0}^{\ell-1} \sum_{i,j=1}^n \left| \sum_{t=\ell}^{T-1} \eta_{i,t+1} u_{j,t-s} \right|^2.$$

Assumption 3.1 justifies the use of Lemma 2 in Wei (1987) which gives $\mathbb{E} \left| \sum_{t=\ell}^{T-1} \eta_{i,t+1} u_{j,t-s} \right|^2 \leq C \sum_{t=\ell}^{T-1} \mathbb{E}(u_{j,t-s}^2) \leq C(T-\ell)$. By Chebyshev's inequality, $\forall \varepsilon > 0$, there exists $\alpha_\varepsilon > 0$ such that

$$\mathbb{P} \left(II_a \geq \alpha_\varepsilon \frac{q}{\sqrt{T}} \right) \leq \frac{1}{\alpha_\varepsilon^2} \frac{T}{q^2} \sum_{\ell=1}^q \mathbb{E} \left\| \frac{1}{T-\ell} \sum_{t=\ell}^{T-1} \boldsymbol{\eta}_{t+1} \mathbf{u}_t(\ell)' \right\|^2 \leq \frac{C}{\alpha_\varepsilon^2} \leq \varepsilon,$$

and thus $II_a = O_p(q/\sqrt{T})$. Furthermore, we deduce that $II_b = O_p(q/\sqrt{T})$ by Lemma 3.2(ii) and Chebyshev's inequality. For II_c , if we write $\widehat{\boldsymbol{\eta}}_{t+1,\ell} - \boldsymbol{\eta}_{t+1,\ell} = [\mathbf{I}_n, -\mathbf{A}(\ell)][\widehat{\mathbf{u}}_{t+1}(\ell+1) - \mathbf{u}_{t+1}(\ell+1)]$, then by Cauchy-Schwarz inequality and Baxter's inequality (leads to $\max_{1 \leq \ell \leq q} \|\mathbf{A}(\ell)\|^2 \leq C$),

$$II_c \leq C \sqrt{\max_{1 \leq \ell \leq q} \frac{1}{T-\ell} \sum_{t=\ell}^{T-1} \|\widehat{\mathbf{u}}_{t+1}(\ell+1) - \mathbf{u}_{t+1}(\ell+1)\|^2} \sqrt{\max_{1 \leq \ell \leq q} \frac{1}{T-\ell} \sum_{t=\ell}^{T-1} \|\mathbf{u}_t(\ell)\|^2} = O_p\left(\frac{q}{\sqrt{T}}\right),$$

where the last step follows from arguments similar to those preceding (3.43). Similarly, $II_d = O_p(q/\sqrt{T})$ and $II_e = O_p(q/T)$. Combining all results, we finally have

$$\max_{1 \leq \ell \leq q} \left\| \frac{1}{T-\ell} \sum_{t=\ell}^{T-1} \widehat{\boldsymbol{\eta}}_{t+1,\ell} \widehat{\mathbf{u}}_t(\ell)' \right\| = O_p\left(\frac{q}{\sqrt{T}}\right). \quad (3.44)$$

By invertibility of $\frac{1}{T-\ell} \sum_{t=\ell}^{T-1} \widehat{\mathbf{u}}_t(\ell) \widehat{\mathbf{u}}_t(\ell)'$, (3.42) and (3.44), $\max_{1 \leq \ell \leq q} \|\widehat{\mathbf{A}}(\ell) - \mathbf{A}(\ell)\| = O_p(q/\sqrt{T})$ follows.

We consider $\max_{1 \leq \ell \leq q} \|\widehat{\mathbf{S}}(\ell) - \mathbf{S}(\ell)\|$ next. By $\widetilde{\boldsymbol{\eta}}_{t+1,\ell} = \widehat{\boldsymbol{\eta}}_{t+1,\ell} - (\widehat{\mathbf{A}}(\ell) - \mathbf{A}(\ell))' \widehat{\mathbf{u}}_t(\ell)$ (see (3.41)), we can write

$$\begin{aligned} \max_{1 \leq \ell \leq q} \|\widehat{\mathbf{S}}(\ell) - \mathbf{S}(\ell)\| &= \max_{1 \leq \ell \leq q} \left\| \frac{1}{T-\ell} \sum_{t=\ell}^{T-1} (\widetilde{\boldsymbol{\eta}}_{t+1,\ell} \widetilde{\boldsymbol{\eta}}_{t+1,\ell}' - \mathbb{E}(\boldsymbol{\eta}_{t+1,\ell} \boldsymbol{\eta}_{t+1,\ell}')) \right\| \\ &= \max_{1 \leq \ell \leq q} \left\| \frac{1}{T-\ell} \sum_{t=\ell}^{T-1} (\widehat{\boldsymbol{\eta}}_{t+1,\ell} \widehat{\boldsymbol{\eta}}_{t+1,\ell}' - \mathbb{E}(\boldsymbol{\eta}_{t+1,\ell} \boldsymbol{\eta}_{t+1,\ell}')) - \frac{1}{T-\ell} \sum_{t=\ell}^{T-1} \widehat{\boldsymbol{\eta}}_{t+1,\ell} \widehat{\mathbf{u}}_t(\ell)' (\widehat{\mathbf{A}}(\ell) - \mathbf{A}(\ell))' \right\|. \end{aligned}$$

Using (3.42) and the invertibility of $\frac{1}{T-\ell} \sum_{t=\ell}^{T-1} \widehat{\mathbf{u}}_t(\ell) \widehat{\mathbf{u}}_t(\ell)'$ for the last equation above, we further obtain an upper bound

$$\begin{aligned} &\max_{1 \leq \ell \leq q} \|\widehat{\mathbf{S}}(\ell) - \mathbf{S}(\ell)\| \\ &\leq \max_{1 \leq \ell \leq q} \left\| \frac{1}{T-\ell} \sum_{t=\ell}^{T-1} (\widehat{\boldsymbol{\eta}}_{t+1,\ell} \widehat{\boldsymbol{\eta}}_{t+1,\ell}' - \mathbb{E}(\boldsymbol{\eta}_{t+1,\ell} \boldsymbol{\eta}_{t+1,\ell}')) \right\| + C \max_{1 \leq \ell \leq q} \left\| \frac{1}{T-\ell} \sum_{t=\ell}^{T-1} \widehat{\boldsymbol{\eta}}_{t+1,\ell} \widehat{\mathbf{u}}_t(\ell)' \right\|^2 \\ &\leq \max_{1 \leq \ell \leq q} \left\| \frac{1}{T-\ell} \sum_{t=\ell}^{T-1} (\widehat{\boldsymbol{\eta}}_{t+1,\ell} \widetilde{\boldsymbol{\eta}}_{t+1,\ell}' - \boldsymbol{\eta}_{t+1,\ell} \boldsymbol{\eta}_{t+1,\ell}') \right\| + O_p\left(\sqrt{\frac{q}{T}}\right) + O_p\left(\frac{q^2}{T}\right), \end{aligned}$$

where the last step above follows from, using Lemma 3.2 (iii) and (3.44),

$$\max_{1 \leq \ell \leq q} \left\| \frac{1}{T-\ell} \sum_{t=\ell}^{T-1} (\boldsymbol{\eta}_{t+1,\ell} \boldsymbol{\eta}_{t+1,\ell}') - \mathbb{E}(\boldsymbol{\eta}_{t+1,\ell} \boldsymbol{\eta}_{t+1,\ell}') \right\| = O_p\left(\sqrt{\frac{q}{T}}\right).$$

By similar arguments as for II_c and II_d above, the first term in the upper bound of $\max_{1 \leq \ell \leq q} \|\widehat{\mathbf{S}}(\ell) - \mathbf{S}(\ell)\|$ above is bounded by

$$\begin{aligned} &\max_{1 \leq \ell \leq q} \frac{1}{T-\ell} \sum_{t=\ell}^{T-1} \|\widehat{\boldsymbol{\eta}}_{t+1,\ell} - \boldsymbol{\eta}_{t+1,\ell}\|^2 \\ &\quad + 2 \sqrt{\max_{1 \leq \ell \leq q} \frac{1}{T-\ell} \sum_{t=\ell}^{T-1} \|\widehat{\boldsymbol{\eta}}_{t+1,\ell} - \boldsymbol{\eta}_{t+1,\ell}\|^2} \sqrt{\max_{1 \leq \ell \leq q} \frac{1}{T-\ell} \sum_{t=\ell}^{T-1} \|\boldsymbol{\eta}_{t+1,\ell}\|^2} = O_p\left(\frac{q}{\sqrt{T}}\right). \end{aligned}$$

Overall, we obtain $\max_{1 \leq \ell \leq q} \|\widehat{\mathbf{S}}(\ell) - \mathbf{S}(\ell)\| = O_p(q/\sqrt{T})$ as well. \blacksquare

The following lemma is a straightforward generalization of the results in Lemma 2 of Cheng et al. (2015) and Propositions 2.1-2.2 of Ing et al. (2016).

Lemma 3.4

Under Assumptions 3.1 and 3.3, we have

$$\|\Sigma_u^{-1}(q) - \Sigma_u^{-1}\| \leq C \frac{1}{\sqrt{q}} \sum_{s=q+1}^{\infty} s \|\mathbf{A}_s\|_{\mathcal{F}}. \quad (3.45)$$

Proof Consider $\|\Sigma_u^{-1}(q) - \Sigma_u^{-1}\|$. A rewriting as in (3.61) shows that $\|\mathbf{M}_u(q) - \mathbf{M}_u\|$ and $\|\mathbf{S}_u^{-1}(q) - \mathbf{S}_u^{-1}\|$ are the two important terms to bound. Hölder's inequality implies

$$\|\mathbf{M}_u(q) - \mathbf{M}_u\| \leq \sqrt{\|\mathbf{M}_u(q) - \mathbf{M}_u\|_1 \|\mathbf{M}_u(q) - \mathbf{M}_u\|_{\infty}}. \quad (3.46)$$

For the matrix 1-norm we are concerned with the maximum absolute column sum. For an arbitrary $(nT \times nT)$ matrix \mathbf{Q} partitioned (block) column-wise, i.e. $\mathbf{Q} = [\mathbf{Q}_1, \mathbf{Q}_2, \dots, \mathbf{Q}_T]$, we have the bound $\|\mathbf{Q}\|_1 = \max_{1 \leq t \leq T} \|\mathbf{Q}_t\|_1 \leq \sqrt{n} \max_{1 \leq t \leq T} \|\mathbf{Q}_t\|_{\mathcal{F}}$. This implies $\|\mathbf{M}_u(q) - \mathbf{M}_u\|_1 \leq \sqrt{n} \max\{I_a, I_b\}$ where

$$I_a = \max_{0 \leq j \leq T-q-2} \left(\sum_{i=0}^{T-q-2-j} \|\mathbf{A}_{q+1+i}(q+1+i+j)\|_{\mathcal{F}}^2 + \sum_{i=\max(1, q+1-j)}^q \|\mathbf{A}_i(i+j) - \mathbf{A}_i(q)\|_{\mathcal{F}}^2 \right)^{1/2},$$

and

$$I_b = \max_{T-q-1 \leq j \leq T-2} \left(\sum_{i=1}^{T-1-j} \|\mathbf{A}_i(i+j) - \mathbf{A}_i(q)\|_{\mathcal{F}}^2 \right)^{1/2}.$$

We will bound the three summations that are encountered in the expressions for I_a and I_b . First, changing the summation index and using c_r -inequality,

$$\begin{aligned} \sum_{i=0}^{T-q-2-j} \|\mathbf{A}_{q+1+i}(q+1+i+j)\|_{\mathcal{F}}^2 &= \sum_{s=q+1}^{T-1-j} \|\mathbf{A}_s(s+j)\|_{\mathcal{F}}^2 \\ &\leq 2 \sum_{s=q+1}^{T-1-j} \|\mathbf{A}_s(s+j) - \mathbf{A}_s\|_{\mathcal{F}}^2 + 2 \sum_{s=q+1}^{T-1-j} \|\mathbf{A}_s\|_{\mathcal{F}}^2 \\ &\leq 2 \sum_{s=q+1}^{T-1-j} s^{-2} \left(s^2 \sum_{k=1}^{s+j} \|\mathbf{A}_k(s+j) - \mathbf{A}_k\|_{\mathcal{F}}^2 \right) + 2 \sum_{s=q+1}^{\infty} \|\mathbf{A}_s\|_{\mathcal{F}}^2. \end{aligned}$$

For convenience, we define

$$\mathcal{K}_q = \left(\sum_{s=q+1}^{\infty} s \|\mathbf{A}_s\|_{\mathcal{F}} \right)^2 \left(\sum_{s=q+1}^{\infty} \frac{1}{s^2} \right). \quad (3.47)$$

For any $j \geq 0$ and $s \geq q + 1$, we have $s^2 \sum_{k=1}^{s+j} \|\mathbf{A}_k(s+j) - \mathbf{A}_k\|_{\mathcal{F}}^2 \leq C s^2 \sum_{k=s+j+1}^{\infty} \|\mathbf{A}_k\|_{\mathcal{F}}^2 \leq C \left(\sum_{k=s+j+1}^{\infty} k \|\mathbf{A}_k\|_{\mathcal{F}} \right)^2$ by the L^2 -Baxter's inequality. The first term in the RHS above is thus bounded by $C\mathcal{K}_q$. Moreover, by Cauchy-Schwartz inequality, the second term can be bounded as $\sum_{s=q+1}^{\infty} \|\mathbf{A}_s\|_{\mathcal{F}}^2 = \sum_{s=q+1}^{\infty} (s^2 \|\mathbf{A}_s\|_{\mathcal{F}}^2) s^{-2} \leq \left(\sum_{s=q+1}^{\infty} s^2 \|\mathbf{A}_s\|_{\mathcal{F}}^2 \right) \left(\sum_{s=q+1}^{\infty} s^{-2} \right) \leq \mathcal{K}_q$. Now the second summation in I_a . We first consider the case $0 \leq j \leq q$, or $\max(1, q+1-j) = q+1-j$, such that

$$\begin{aligned} \sum_{i=\max(1, q+1-j)}^q \|\mathbf{A}_i(i+j) - \mathbf{A}_i(q)\|_{\mathcal{F}}^2 &= \sum_{i=q+1-j}^q \|\mathbf{A}_i(i+j) - \mathbf{A}_i(q)\|_{\mathcal{F}}^2 \\ &\leq 2 \sum_{i=q+1-j}^q \|\mathbf{A}_i(i+j) - \mathbf{A}_i\|_{\mathcal{F}}^2 + 2 \sum_{i=q+1-j}^q \|\mathbf{A}_i(q) - \mathbf{A}_i\|_{\mathcal{F}}^2 \\ &\leq 2 \sum_{s=q+1}^{q+j} \|\mathbf{A}_{s-j}(s) - \mathbf{A}_{s-j}\|_{\mathcal{F}}^2 + 2 \sum_{i=1}^q \|\mathbf{A}_i(q) - \mathbf{A}_i\|_{\mathcal{F}}^2 \\ &\leq C\mathcal{K}_q \end{aligned}$$

using arguments detailed before. This upper bound remains valid for $q+1 \leq j \leq T-q-2$. It is likewise straightforward to derive $\sum_{i=1}^{T-1-j} \|\mathbf{A}_i(i+j) - \mathbf{A}_i(q)\|_{\mathcal{F}}^2 \leq C\mathcal{K}_q$. Collecting all the results, we have $I_a \leq C\sqrt{\mathcal{K}_q}$, $I_b \leq C\sqrt{\mathcal{K}_q}$, and thus $\|\mathbf{F}_u(q) - \mathbf{F}_u\|_1 \leq C\sqrt{\mathcal{K}_q}$.

For $\|\mathbf{M}_u(q) - \mathbf{M}_u\|_{\infty}$, we are bounding the maximum absolute row sums. For an arbitrary $(nT \times nT)$ matrix \mathbf{Q} partitioned as $\mathbf{Q} = [\mathbf{Q}'_1, \mathbf{Q}'_2, \dots, \mathbf{Q}'_T]'$, we have $\|\mathbf{Q}\|_{\infty} = \max_{1 \leq t \leq T} \|\mathbf{Q}_t\|_{\infty} \leq \sqrt{n} \max_{1 \leq t \leq T} \|\mathbf{Q}_t\|_{\mathcal{F}}$, such that

$$\|\mathbf{M}_u(q) - \mathbf{M}_u\|_{\infty} \leq \sqrt{n} \max_{q+1 \leq m \leq T-1} \left\{ \sum_{j=q+1}^m \|\mathbf{A}_j(m)\|_{\mathcal{F}}^2 + \sum_{j=1}^q \|\mathbf{A}_j(q) - \mathbf{A}_j(m)\|_{\mathcal{F}}^2 \right\}^{1/2},$$

where $\sum_{j=q+1}^m \|\mathbf{A}_j(m)\|_{\mathcal{F}}^2 \leq C\mathcal{K}_q$ and $\sum_{j=1}^q \|\mathbf{A}_j(q) - \mathbf{A}_j(m)\|_{\mathcal{F}}^2 \leq C\mathcal{K}_q$, for any $q+1 \leq m \leq T-1$, using the L^2 -Baxter's inequality and the previous upper bound on $\sum_{s=q+1}^{\infty} \|\mathbf{A}_s\|_{\mathcal{F}}^2$. We conclude that $\|\mathbf{M}_u(q) - \mathbf{M}_u\|_{\infty} \leq C\sqrt{\mathcal{K}_q}$. Together with our previous result, we obtain $\|\mathbf{M}_u(q) - \mathbf{M}_u\| \leq C\sqrt{\mathcal{K}_q}$ from (3.46).

From $\|\mathbf{S}_u^{-1}(q) - \mathbf{S}_u^{-1}\| \leq \|\mathbf{S}_u(q) - \mathbf{S}_u\| \|\mathbf{S}_u^{-1}\| \|\mathbf{S}_u^{-1}(q)\|$ we see that it suffices to inspect $\|\mathbf{S}_u(q) - \mathbf{S}_u\|$ (the other norms are bounded). Exploiting the fact that both $\mathbf{S}_u(q)$

and \mathbf{S}_u are block-diagonal, we have $\|\mathbf{S}_u(q) - \mathbf{S}_u\| = \max_{q+1 \leq k \leq T-1} \|\mathbf{S}(q) - \mathbf{S}(k)\| \leq 2 \max_{q \leq k \leq T-1} \|\mathbf{S}(k) - \Sigma_{\eta\eta}\|$. Let $\mathbf{A}_j(\ell) = \mathbf{O}$ for $j > \ell$, and recall the definition of $\eta_{t+1,\ell}$ in (3.35). We find, for any $k \geq q$,

$$\begin{aligned} \|\mathbf{S}(k) - \Sigma_{\eta\eta}\| &= \|\mathbb{E}(\eta_{t+1,k} - \eta_{t+1}) (\eta_{t+1,k} - \eta_{t+1})'\| \\ &\leq C \sum_{s=1}^{\infty} \|\mathbf{A}_s(k) - \mathbf{A}_s\|^2 \leq C \sum_{s=q+1}^{\infty} \|\mathbf{A}_s\|_{\mathcal{F}}^2 \leq C\mathcal{K}_q. \end{aligned} \quad (3.48)$$

We thereby obtain $\|\mathbf{S}_u^{-1}(q) - \mathbf{S}_u^{-1}\| \leq C\mathcal{K}_q$. Together with the bound on $\|\mathbf{M}_u(q) - \mathbf{M}_u\|$, we deduce

$$\|\Sigma_u^{-1}(q) - \Sigma_u^{-1}\| \leq C\sqrt{\mathcal{K}_q} \leq C \frac{1}{\sqrt{q}} \sum_{s=q+1}^{\infty} s \|\mathbf{A}_s\|_{\mathcal{F}}.$$

The proof is complete. ■

Appendix 3.B Proofs of Main Theorems

Lemma 3.5

Let $\mathcal{A}_q(L) = \mathbf{I}_n - \sum_{j=1}^q \mathbf{A}_j(q)L^j$ denote the lag polynomial implied by the coefficient matrices in (3.3). By the Beveridge-Nelson (BN) decomposition, we also have $\mathcal{A}_q(L) = \mathcal{A}_q(1) + (1-L)\tilde{\mathcal{A}}_q(L)$ where $\tilde{\mathcal{A}}_q(L) = \sum_{j=1}^q \tilde{\mathbf{A}}_j(q)L^{j-1}$ with $\tilde{\mathbf{A}}_j(q) = \sum_{i=j}^q \mathbf{A}_i(q)$. If Assumption 3.1 holds, then

- (a) $\mathcal{A}_q(1) = \mathcal{A}(1) + \mathbf{O}\left(\sum_{j=q+1}^{\infty} j^{1/2} \|\mathbf{A}_j\|_{\mathcal{F}}\right)$
- (b) There exists a $q^* > 0$ such that $\sum_{j=1}^q \|\tilde{\mathbf{A}}_j(q)\|_{\mathcal{F}} < \infty$ for all $q > q^*$

Proof (a) By Cauchy-Schwartz, we have $\|\mathcal{A}_q(1) - \mathcal{A}(1)\|_{\mathcal{F}} \leq \sum_{j=1}^q \|\mathbf{A}_j(q) - \mathbf{A}_j\|_{\mathcal{F}} + \sum_{j=q+1}^{\infty} \|\mathbf{A}_j\|_{\mathcal{F}} \leq (q \sum_{j=1}^q \|\mathbf{A}_j(q) - \mathbf{A}_j\|_{\mathcal{F}}^2)^{1/2} + \sum_{j=q+1}^{\infty} \|\mathbf{A}_j\|_{\mathcal{F}}$ and subsequently use Lemma 3.1. (b) Using Baxter's inequality in Theorem 6.6.12 in Hannan and Deistler (2012) (also see their Remark 3) for the final inequality, we derive $\sum_{j=1}^q \|\tilde{\mathbf{A}}_j(q)\|_{\mathcal{F}} \leq \sum_{j=1}^q j \|\mathbf{A}_j(q)\|_{\mathcal{F}} \leq \sum_{j=1}^q j \|\mathbf{A}_j(q) - \mathbf{A}_j\|_{\mathcal{F}} + \sum_{j=1}^q j \|\mathbf{A}_j\|_{\mathcal{F}} \leq C \sum_{j=1}^q j \|\mathbf{A}_j\|_{\mathcal{F}} \leq C$. ■

Proof of Theorem 3.1 The premultiplication by $\mathbf{M}_u(q)$ applies a linear filter whereas $\mathbf{S}_u^{-1}(q)$ implies weighting. Since the behaviour of the first $q \ll T$ elements does not

affect the asymptotic results, we take $\mathbf{Z}_t = \mathbf{u}_t = \mathbf{O}$ for $t \leq 0$ and for all $t = 1, 2, \dots$, we apply the transformations implied by $\mathcal{A}_q(L)$ and $S^{-1}(q)$.¹² Consequently, we have

$$\begin{aligned} \mathbf{G}_T^{-1} (\widehat{\beta}_{GLS} - \beta) &= \left[\sum_{t=1}^T (\mathcal{A}_q(L) \mathbf{Z}_t' \mathbf{G}_T)' S^{-1}(q) (\mathcal{A}_q(L) \mathbf{Z}_t' \mathbf{G}_T) \right]^{-1} \\ &\quad \times \sum_{t=1}^T (\mathcal{A}_q(L) \mathbf{Z}_t' \mathbf{G}_T)' S^{-1}(q) (\mathcal{A}_q(L) \mathbf{u}_t) + o_p(1). \end{aligned} \quad (3.49)$$

Using the BN decomposition, Lemma 3.5, we first find

$$\begin{aligned} \mathcal{A}_q(L) \mathbf{Z}_t' \mathbf{G}_T \\ = \mathcal{A}_q(1) \mathbf{Z}_t' \mathbf{G}_T + \sum_{j=1}^q \widetilde{\mathbf{A}}_j(q) \Delta \mathbf{Z}_{t-j+1}' \mathbf{G}_T = \mathcal{A}(1) \mathbf{Z}_t' \mathbf{G}_T + O_p(1/\sqrt{qT}). \end{aligned} \quad (3.50)$$

Note that $\left\| \sum_{j=1}^q \widetilde{\mathbf{A}}_j(q) \Delta \mathbf{Z}_{t-j+1}' \mathbf{G}_T \right\| \leq \sum_{j=1}^q \|\widetilde{\mathbf{A}}_j(q)\| \|\Delta \mathbf{Z}_{t-j+1}' \mathbf{G}_T\|$ and that for any t ,

$$\|\Delta \mathbf{Z}_t' \mathbf{G}_T\| = \max_{1 \leq i \leq n} \|\mathbf{G}_{i,T} \Delta \mathbf{z}_{it}\| = \max_{1 \leq i \leq n} \left(\|\mathbf{G}_{d_{i,T}} \Delta \mathbf{d}_{it}\|^2 + \|\mathbf{G}_{s_{i,T}} \Delta \mathbf{s}_{it}\|^2 \right)^{1/2}. \quad (3.51)$$

The vector $\mathbf{G}_{d_{i,T}} \Delta \mathbf{d}_{it}$ typically contains elements $T^{-(k+\frac{1}{2})} [t^k - (t-1)^k]$ where $k = 0, 1, \dots, q_i$. By the inequality $(a+b)^n \leq a^n + nb(a+b)^{n-1}$, for $a, b \geq 0$, $n \in \mathbb{N}$, we obtain $0 \leq t^k - (t-1)^k \leq kt^{k-1}$, and thus $T^{-(k+\frac{1}{2})} [t^k - (t-1)^k] \leq q_i T^{-3/2} \leq CT^{-3/2}$. As a result, $\|\mathbf{G}_{d_{i,T}} \Delta \mathbf{d}_{it}\|^2 \leq CT^{-3}$. The vector $\mathbf{G}_{s_{i,T}} \Delta \mathbf{s}_{it}$ typically contains elements $T^{-(k+1)/2} (x_{it}^k - x_{it-1}^k)$, where $k = 1, \dots, p_i$. The binomial expansion implies $x_{it}^k - x_{it-1}^k = \sum_{m=0}^{k-1} \binom{k}{m} x_{it-1}^m v_{it}^{k-m} = O_p(T^{-(k-1)/2})$, and thus $T^{-(k+1)/2} (x_{it}^k - x_{it-1}^k) = O_p(T^{-1})$. It further implies that $\|\mathbf{G}_{s_{i,T}} \Delta \mathbf{s}_{it}\|^2 = O_p(T^{-2})$. Combining $\|\Delta \mathbf{Z}_t' \mathbf{G}_T\| = O_p(T^{-1})$ with Lemma 3.5(b) we establish the second equality in (3.50). Now (3.50) follows from Lemma 3.5(a) and $\sum_{j=q+1}^{\infty} j^{1/2} \|\mathbf{A}_j\|_{\mathcal{F}} = o(q^{-1/2})$. Using (3.50) and $\|\mathbf{S}(q) - \Sigma_{\eta\eta}\| \rightarrow 0$ (see (3.48)), we obtain

$$\begin{aligned} \sum_{t=1}^T (\mathcal{A}_q(L) \mathbf{Z}_t' \mathbf{G}_T)' S^{-1}(q) (\mathcal{A}_q(L) \mathbf{Z}_t' \mathbf{G}_T) \\ = \sum_{t=1}^T \mathbf{G}_T \mathbf{Z}_t' \Omega_{uu}^{-1} \mathbf{Z}_t' \mathbf{G}_T + O_p(q^{-1/2}) \Rightarrow \int_0^1 \mathbf{J}(r) \Omega_{uu}^{-1} \mathbf{J}(r)' dr. \end{aligned} \quad (3.52)$$

¹²The same argumentation is used in Phillips and Park (1988). The modification to obtain a rigorous proof is straightforward.

We rewrite the second part in (3.49) as

$$\begin{aligned} \sum_{t=1}^T \left(\mathcal{A}_q(L) \mathbf{Z}_t' \mathbf{G}_T \right)' \mathbf{S}^{-1}(q) \left(\mathcal{A}_q(L) \mathbf{u}_t \right) &= \sum_{t=1}^T \left(\mathcal{A}_q(L) \mathbf{Z}_t' \mathbf{G}_T \right)' \mathbf{S}^{-1}(q) \boldsymbol{\eta}_t \\ &+ \sum_{t=1}^T \left(\mathcal{A}_q(L) \mathbf{Z}_t' \mathbf{G}_T \right)' \mathbf{S}^{-1}(q) \left(\mathcal{A}_q(L) \mathbf{u}_t - \boldsymbol{\eta}_t \right) =: I + II, \end{aligned} \quad (3.53)$$

and we will repeatedly use the identity

$$\begin{aligned} \sum_t \mathbf{J}_t \mathbf{D} \mathbf{e}_t &= \sum_t (\mathbf{e}_t' \otimes \mathbf{J}_t) \text{vec}(\mathbf{D}) \\ &= \left[\sum_t \mathbf{J}_t \mathbf{e}_{1t}, \sum_t \mathbf{J}_t \mathbf{e}_{2t}, \dots, \sum_t \mathbf{J}_t \mathbf{e}_{nt} \right] \text{vec}(\mathbf{D}) \end{aligned} \quad (3.54)$$

for any matrices $\mathbf{J}_t \in \mathbb{R}^{d \times n}$, $\mathbf{D} \in \mathbb{R}^{n \times n}$ and $\mathbf{e}_t = [e_{1t}, \dots, e_{nt}]' \in \mathbb{R}^{n \times 1}$. Using this identity, we have

$$I = \left[\sum_{t=1}^T \left(\mathcal{A}_q(L) \mathbf{Z}_t' \mathbf{G}_T \right)' \eta_{1t}, \dots, \sum_{t=1}^T \left(\mathcal{A}_q(L) \mathbf{Z}_t' \mathbf{G}_T \right)' \eta_{nt} \right] \text{vec} \left(\boldsymbol{\Sigma}_{\eta\eta}^{-1} + o(1) \right), \quad (3.55)$$

where η_{it} is the i^{th} entry of $\boldsymbol{\eta}_t$. For $1 \leq i \leq n$, we subsequently use the BN decomposition to obtain:

$$\begin{aligned} \sum_{t=1}^T \left(\mathcal{A}_q(L) \mathbf{Z}_t' \mathbf{G}_T \right)' \eta_{it} &= \sum_{t=1}^T \mathbf{G}_T \mathbf{Z}_t \eta_{it} \mathcal{A}_q(1)' \\ &+ \sum_{t=1}^T \mathbf{G}_T \Delta \mathbf{Z}_t \eta_{it} \tilde{\mathbf{A}}_1(q)' + \sum_{j=2}^q \sum_{t=1}^T \mathbf{G}_T \Delta \mathbf{Z}_{t-j+1} \eta_{it} \tilde{\mathbf{A}}_j(q)'. \end{aligned} \quad (3.56)$$

We have $\sum_{t=1}^T \mathbf{G}_T \mathbf{Z}_t \eta_{it} = \text{diag} \left[\sum_{t=1}^T \mathbf{G}_{1,T} \mathbf{z}_{1t}, \dots, \sum_{t=1}^T \mathbf{G}_{n,T} \mathbf{z}_{nt} \right] \eta_{it}$ by definition, where the limiting distribution of each block follows from Proposition 1 of Wagner and Hong (2016). More specifically, the k^{th} block will converge to a stochastic integral and a second order bias term which is proportional to $\boldsymbol{\Sigma}_{\epsilon_k \eta_i} := \mathbb{E}(\epsilon_{kt} \eta_{it})$ (the $(k, i)^{\text{th}}$ element of $\boldsymbol{\Sigma}_{\epsilon\eta}$), $1 \leq k, i \leq n$, and thus

$$\sum_{t=1}^T \mathbf{G}_T \mathbf{Z}_t \eta_{it} \mathcal{A}_q(1)' \Rightarrow \left(\int_0^1 \mathbf{J}(r) d\mathbf{B}_{\eta_i}(r) + \mathbf{B}_i \right) \mathcal{A}(1)', \quad (3.57)$$

where $B_i := \text{diag} \left[\sum_{\epsilon_i \eta_i} \mathbf{b}_1, \dots, \sum_{\epsilon_i \eta_i} \mathbf{b}_n \right]$.

Now we consider the second term in (3.56). As $\sum_{t=1}^T \mathbf{G}_T \Delta \mathbf{Z}_t \eta_{it} \tilde{\mathbf{A}}_1(q)' = \text{diag} \left[\sum_{t=1}^T \mathbf{G}_{1,T} \Delta \mathbf{z}_{1t}, \dots, \sum_{t=1}^T \mathbf{G}_{n,T} \Delta \mathbf{z}_{nt} \right] \eta_{it} \left(\sum_{j=1}^{\infty} \mathbf{A}_j + o(1) \right)'$, we again consider the limiting distributions block-wise. Every element in the blocks will rely on one of the following three results. (1) As derived below (3.51), we have $|T^{-(j+\frac{1}{2})} \sum_{t=1}^T [t^j - (t-1)^j] \eta_{it}| \leq CT^{-3/2} \sum_{t=1}^T |\eta_{it}| = o_p(1)$. (2) By Assumption 3.1, for any $1 \leq k, i \leq n$, v_{kt} and η_{it} are Near Epoch Dependent in L_4 -norm on $\{[\eta'_t, \epsilon'_t]'\}_{t \in \mathbb{Z}}$ of size -1 and arbitrary size, respectively. A small variation on Theorem 17.9 from Davidson (1994) shows that $\{v_{it} \eta_{it}\}$ are L_2 -NED of size -1 . The i.i.d. assumption on $\{[\eta'_t, \epsilon'_t]'\}$ allows for a LLN for the sequence $\{v_{it} \eta_{it}\}$, see e.g. Theorem 20.21 of Davidson (1994), implying $T^{-1} \sum_{t=1}^T \Delta x_{kt} \eta_{it} = T^{-1} \sum_{t=1}^T v_{kt} \eta_{it} \rightarrow_p \sum_{\epsilon_k \eta_i}$. (3) $T^{-(j+1)/2} \sum_{t=1}^T \Delta x_{kt}^j \eta_{it} \Rightarrow j \sum_{\epsilon_k \eta_i} \int_0^1 B_{v_k}^{j-1}(r) dr$, where $j \geq 2$. The specific reason is as follows. By the binomial expansion (below (3.51)), we have

$$\begin{aligned} T^{-(j+1)/2} \sum_{t=1}^T \Delta x_{kt}^j \eta_{it} &= T^{-(j+1)/2} \sum_{m=0}^{j-1} \binom{j}{m} \sum_{t=1}^T x_{kt-m}^m v_{kt}^{j-m} \eta_{it} \\ &= j T^{-(j+1)/2} \sum_{t=1}^T x_{kt-1}^{j-1} v_{kt} \eta_{it} + o_p(1) \\ &= j \sum_{\epsilon_k \eta_i} \frac{1}{T} \sum_{t=1}^{T-1} \left(\frac{x_{kt}}{\sqrt{T}} \right)^{j-1} + \frac{j}{\sqrt{T}} \frac{1}{\sqrt{T}} \sum_{t=1}^{T-1} \left(\frac{x_{kt}}{\sqrt{T}} \right)^{j-1} (v_{kt} \eta_{it} - \sum_{\epsilon_k \eta_i}) + o_p(1) \\ &\Rightarrow j \sum_{\epsilon_k \eta_i} \int_0^1 B_{v_k}^{j-1}(r) dr, \end{aligned}$$

where $\frac{1}{\sqrt{T}} \sum_{t=1}^{T-1} \left(\frac{x_{kt}}{\sqrt{T}} \right)^{j-1} (v_{kt} \eta_{it} - \sum_{\epsilon_k \eta_i}) = O_p(1)$. To see this, we refer to de Jong (2002). The moment and NED conditions in his Assumption 1 are satisfied. Moreover, since $F(x) = x^{j-1}$ is homogeneous of degree $j-1$, his Assumption 2 holds as well. The desired result now follows from Theorem 1 in de Jong (2002). Combining these results, we obtain

$$\sum_{t=1}^T \mathbf{G}_T \Delta \mathbf{Z}_t \eta_{it} \tilde{\mathbf{A}}_1(q)' \Rightarrow \mathbf{B}_i \left(\sum_{j=1}^{\infty} \mathbf{A}_j \right)'. \quad (3.58)$$

The final term in (3.56) is bounded by $\sum_{j=2}^q \|\mathbf{A}_j(q)\| \|\sum_{t=1}^T \mathbf{G}_T \Delta \mathbf{Z}_{t-j+1} \eta_{it}\|$. Using similar arguments above, we conclude that $\sum_{t=1}^T \mathbf{G}_T \Delta \mathbf{Z}_{t-j+1} \eta_{it} = o_p(1)$ (lags of $\Delta \mathbf{Z}_t$ will lead to $\mathbb{E}(v_{kt-j} \eta_{it}) = 0$ for any $j > 0$). Hence, $\sum_{j=2}^q \sum_{t=1}^T \mathbf{G}_T \Delta \mathbf{Z}_{t-j+1} \eta_{it} \tilde{\mathbf{A}}_j(q)' =$

$o_p(1)$. Combining (3.56), (3.57) and (3.58), we have

$$\sum_{t=1}^T (\mathcal{A}_q(L) \mathbf{Z}_t' \mathbf{G}_T)' \eta_{it} \Rightarrow \int_0^1 \mathbf{J}(r) d\mathbf{B}_{\eta_i}(r) \mathcal{A}(1)' + \mathbf{B}_i. \quad (3.59)$$

Note that $\text{vec}(\Sigma_{\eta\eta}^{-1}) = \begin{bmatrix} \text{col}_1(\Sigma_{\eta\eta}^{-1}) \\ \vdots \\ \text{col}_n(\Sigma_{\eta\eta}^{-1}) \end{bmatrix}$. Inserting (3.59) into (3.55), for the term I in (3.53), we have

$$\begin{aligned} I &\Rightarrow \int_0^1 (d\mathbf{B}_{\eta}(r)' \otimes \mathbf{J}(r) \mathcal{A}(1)') \text{vec}(\Sigma_{\eta\eta}^{-1}) + [\mathbf{B}_1, \dots, \mathbf{B}_n] \text{vec}(\Sigma_{\eta\eta}^{-1}) \\ &= \int_0^1 \mathbf{J}(r) \mathcal{A}(1)' \Sigma_{\eta\eta}^{-1} d\mathbf{B}_{\eta}(r) + \sum_{i=1}^n \mathbf{B}_i \text{col}_i(\Sigma_{\eta\eta}^{-1}) \\ &= \int_0^1 \mathbf{J}(r) \Omega_{uu}^{-1} d\mathbf{B}_u(r) + \mathcal{B}_{\epsilon\eta} \end{aligned} \quad (3.60)$$

where the symmetry property of $\Sigma_{\eta\eta}^{-1}$ is used in the final step.

Consider the term II in (3.53). If we define

$$\mathbf{u}_t^* = [u_{1t}^*, \dots, u_{nt}^*] := \mathcal{A}_q(L) \mathbf{u}_t - \boldsymbol{\eta}_t = - \sum_{j=1}^{\infty} (\mathbf{A}_j(q) - \mathbf{A}_j) \mathbf{u}_{t-j},$$

where $\mathbf{A}_j(q) = \mathbf{O}$ for $j > q$, and then apply (3.54), we have

$$II = \left[\sum_{t=1}^T (\mathcal{A}_q(L) \mathbf{Z}_t' \mathbf{G}_T)' u_{1t}^*, \dots, \sum_{t=1}^T (\mathcal{A}_q(L) \mathbf{Z}_t' \mathbf{G}_T)' u_{nt}^* \right] \text{vec}(\Sigma_{\eta\eta}^{-1} + o(1)).$$

For any block $i = 1, \dots, n$, by the BN decomposition (3.50), we have

$$\begin{aligned} \left\| \sum_{t=1}^T (\mathcal{A}_q(L) \mathbf{Z}_t' \mathbf{G}_T)' u_{it}^* \right\| &= \left\| \sum_{t=1}^T \mathbf{G}_T \mathbf{Z}_t u_{it}^* \mathcal{A}_q(1)' + O_p(T^{-1}) \sum_{t=1}^T u_{it}^* \right\| \\ &\leq C \max_{1 \leq k \leq n} \left\| \sum_{t=1}^T \mathbf{G}_{k,T} \mathbf{z}_{kt} u_{it}^* \right\| + O_p(T^{-1}) \sum_{j=1}^{\infty} \left\| \text{row}_i(\mathbf{A}_j(q) - \mathbf{A}_j) \right\| \left\| \sum_{t=1}^T \mathbf{u}_{t-j} \right\| \\ &= O_p \left(\sum_{j=q+1}^{\infty} j^{1/2} \|\mathbf{A}_j\|_{\mathcal{F}} \right) = o_p(1). \end{aligned}$$

It implies $II = o_p(1)$. The theorem follows from (3.49), (3.52), and (3.60). \blacksquare

Proof of Theorem 3.2 We start with the estimation error $\widehat{\Sigma}_u^{-1}(q) - \Sigma_u^{-1}(q)$. Repeated addition and subtraction yields

$$\begin{aligned} \|\widehat{\Sigma}_u^{-1}(q) - \Sigma_u^{-1}(q)\| &\leq \|\widehat{\mathcal{M}}_u(q) - \mathcal{M}_u(q)\| \|\widehat{\mathcal{S}}_u^{-1}(q)\| \|\widehat{\mathcal{M}}_u(q)\| \\ &\quad + \|\mathcal{M}_u(q)\| \|\widehat{\mathcal{S}}_u^{-1}(q) - \mathcal{S}_u^{-1}(q)\| \|\widehat{\mathcal{M}}_u(q)\| \\ &\quad + \|\mathcal{M}_u(q)\| \|\mathcal{S}_u^{-1}(q)\| \|\widehat{\mathcal{M}}_u(q) - \mathcal{M}_u(q)\|. \end{aligned} \quad (3.61)$$

We will only consider the terms $\|\widehat{\mathcal{M}}_u(q) - \mathcal{M}_u(q)\|$ and $\|\widehat{\mathcal{S}}_u^{-1}(q) - \mathcal{S}_u^{-1}(q)\|$. It is not hard to derive that the remaining terms are bounded in probability. Define $\mathcal{G} = \widehat{\mathcal{M}}_u(q) - \mathcal{M}_u(q)$ and denote its $(n \times n)$ subblocks by \mathcal{G}_{ij} , $1 \leq i, j \leq T$. This matrix \mathcal{G} is banded in such a way that there are at most $2q - 1$ nonzero block in the block-columns of $\mathcal{G}\mathcal{G}'$. Using this observation and various norm properties, we find

$$\begin{aligned} \|\mathcal{G}\|^2 &\leq \|\mathcal{G}\mathcal{G}'\|_1 \leq \max_{1 \leq j \leq T} \sum_{i=1}^T \|(\mathcal{G}\mathcal{G}')_{ij}\|_1 \leq n(2q - 1) \max_{1 \leq j \leq T} \sum_{i=1}^T \|\mathcal{G}_{ji}\|_1^2 \\ &= n(2q - 1) \max_{1 \leq \ell \leq q} \sum_{s=1}^{\ell} \|\widehat{\mathcal{A}}_s(\ell) - \mathcal{A}_s(\ell)\|_1^2 \\ &\leq n^3(2q - 1) \max_{1 \leq \ell \leq q} \|\widehat{\mathcal{A}}(\ell) - \mathcal{A}(\ell)\|^2 \\ &= O_p\left(\frac{q^3}{T}\right), \end{aligned}$$

where the final step follows from Lemma 3.3.¹³ We obtain $\|\widehat{\mathcal{M}}_u(q) - \mathcal{M}_u(q)\| = O_p\left(\sqrt{q^3/T}\right)$.

The difference $\widehat{\mathcal{S}}_u^{-1}(q) - \mathcal{S}_u^{-1}(q)$ forms a symmetric and block diagonal matrix, hence $\|\widehat{\mathcal{S}}_u(q) - \mathcal{S}_u(q)\| = \max\{\|\widehat{\mathcal{S}}(0) - \mathcal{S}(0)\|, \max_{1 \leq \ell \leq q} \|\widehat{\mathcal{S}}(\ell) - \mathcal{S}(\ell)\|\}$. By Assumption 3.2,

$$\begin{aligned} \|\widehat{\mathcal{S}}(0) - \mathcal{S}(0)\| &\leq \frac{1}{T} \|\widehat{\mathbf{u}} - \mathbf{u}\|^2 \\ &\quad + 2 \left\| \frac{1}{T} \sum_{t=1}^T (\widehat{\mathbf{u}}_t - \mathbf{u}_t) \mathbf{u}_t' \right\| + \left\| \frac{1}{T} \sum_{t=1}^T \mathbf{u}_t \mathbf{u}_t' - \mathbb{E}(\mathbf{u}_t \mathbf{u}_t') \right\| = O_p(T^{-1/2}). \end{aligned}$$

¹³More specifically, for any matrix \mathbf{Q} we have $\|\mathbf{Q}\|^2 \leq \|\mathbf{Q}\mathbf{Q}'\|_1$. Moreover, if \mathbf{Q} is an $(n \times n)$ matrix, then also $\|\mathbf{Q}\|_1 \leq \sqrt{n} \|\mathbf{Q}\|_{\mathcal{F}} \leq n \|\mathbf{Q}\|$.

Applying Lemma 3.3, we see

$$\left\| \widehat{\mathcal{S}}_u^{-1}(q) - \mathcal{S}_u^{-1}(q) \right\| \leq \left\| \widehat{\mathcal{S}}_u(q) - \mathcal{S}_u(q) \right\| \left\| \widehat{\mathcal{S}}_u^{-1}(q) \right\| \left\| \mathcal{S}_u^{-1}(q) \right\| = O_p \left(\frac{q}{\sqrt{T}} \right).$$

Overall, recalling (3.61), a bound on the estimation error can be given by $\left\| \widehat{\Sigma}_u^{-1}(q) - \Sigma_u^{-1}(q) \right\| = O_p \left(\sqrt{q^3/T} \right)$. The bound on the truncation error, $\left\| \Sigma_u^{-1}(q) - \Sigma_u^{-1} \right\| \leq C \frac{1}{\sqrt{q}} \sum_{s=q+1}^{\infty} s \|\mathcal{A}_s\|_{\mathcal{F}}$ is given in Lemma 3.4. ■

Proof of Theorem 3.3 (a)-(b) See the proof of Proposition 1 in Wagner et al. (2020). **(c)** Since the residuals $\{\widehat{u}_t\}$ are obtained by first stage OLS, we get $\|\widehat{u} - u\|^2 \leq \|\mathcal{G}_T^{-1}(\widehat{\beta}_{OLS} - \beta)\|^2 \|\mathcal{G}_T \mathbf{Z}' \mathbf{Z} \mathcal{G}_T\| = O_p(1)$. Assumption 3.2 is thus satisfied and we can rely on the results in Theorem 3.2. From (3.24), the definition of the FM-GLS estimator, we have

$$\begin{aligned} \mathcal{G}_T^{-1} \left(\widehat{\beta}_{FGLS}^+ - \beta \right) &= \left(\mathcal{G}_T \mathbf{Z}' \widehat{\Sigma}_u^{-1}(q) \mathbf{Z} \mathcal{G}_T \right)^{-1} \\ &\quad \times \left[\mathcal{G}_T \mathbf{Z}' \widehat{\Sigma}_u^{-1}(q) u - \mathcal{G}_T \mathbf{Z}' (\mathbf{I}_T \otimes \widehat{\Omega}_{uu}^{-1} \widehat{\Omega}_{uv} \widehat{\Omega}_{vv}^{-1}) v - \mathcal{G}_T \widehat{\mathcal{B}}^+ \right]. \end{aligned}$$

Given Theorem 3.2, we have $\mathcal{G}_T \mathbf{Z}' \widehat{\Sigma}_u^{-1}(q) \mathbf{Z} \mathcal{G}_T = \mathcal{G}_T \mathbf{Z}' \Sigma_u^{-1}(q) \mathbf{Z} \mathcal{G}_T + o_p(1)$ and it converges weakly to the expression in (3.52).

To continue, we define $\widehat{\mathcal{A}}_q(L) = \mathbf{I}_n - \sum_{j=1}^q \widehat{\mathcal{A}}_j(q) L^j$ and its BN decomposition $\widehat{\mathcal{A}}_q(L) = \widehat{\mathcal{A}}_q(1) + (1-L)\mathcal{A}_q^*(L)$ through $\mathcal{A}_q^*(L) = \sum_{j=1}^q \mathcal{A}_j^*(q) L^{j-1}$ with $\mathcal{A}_j^*(q) = \sum_{i=j}^q \widehat{\mathcal{A}}_i(q)$. $\widehat{\mathcal{A}}_q(1) = \mathcal{A}_q(1) + o_p(1)$ and $\sum_{j=1}^q \|\mathcal{A}_j^*(q)\|_{\mathcal{F}} \leq \sum_{j=1}^q \|\widehat{\mathcal{A}}_j(q)\|_{\mathcal{F}} + o_p(q)$ are obtained from the following two results: (1) $\|\widehat{\mathcal{A}}_q(1) - \mathcal{A}_q(1)\|_{\mathcal{F}} \leq \sum_{j=1}^q \|\widehat{\mathcal{A}}_j(q) - \mathcal{A}_j(q)\|_{\mathcal{F}} \leq C \sqrt{q} \|\widehat{\mathcal{A}}(q) - \mathcal{A}(q)\| = O_p \left(\frac{q^{3/2}}{T^{1/2}} \right) = o_p(1)$, where the last step follows from Lemma 3.3, and (2) $\sum_{j=1}^q \|\mathcal{A}_j^*(q) - \widehat{\mathcal{A}}_j(q)\|_{\mathcal{F}} \leq q \sum_{j=1}^q \|\widehat{\mathcal{A}}_j(q) - \mathcal{A}_j(q)\|_{\mathcal{F}} = o_p(q)$. Using the BN decomposition of $\widehat{\mathcal{A}}_q(L)$ and similar steps as those below (3.53), we have

$$\begin{aligned} &\mathcal{G}_T \mathbf{Z}' \widehat{\Sigma}_u^{-1}(q) u \\ &= \sum_{t=1}^T \left(\widehat{\mathcal{A}}_q(L) \mathbf{Z}'_t \mathcal{G}_T \right)' \widehat{\mathcal{S}}^{-1}(q) \left(\widehat{\mathcal{A}}_q(L) u_t \right) + o_p(1) \Rightarrow \int_0^1 \mathbf{J}(r) \Omega_{uu}^{-1} d\mathcal{B}_u(r) + \mathcal{B}_{\varepsilon\eta}, \end{aligned}$$

where $\widehat{\mathcal{S}}(q) = \mathcal{S}(q) + o_p(1)$ given in Lemma 3.3. Using the identity (3.54), it is not

hard to deduce

$$G_T Z' (I_T \otimes \widehat{\Omega}_{uu}^{-1} \widehat{\Omega}_{uv} \widehat{\Omega}_{vv}^{-1}) v \\ \Rightarrow \int_0^1 J(r) \Omega_{uu}^{-1} \Omega_{uv} \Omega_{vv}^{-1} dB_v(r) + \underbrace{\begin{bmatrix} \text{row}_1(\Delta_{vv}) \text{col}_1(\Omega_{vv}^{-1} \Omega_{vu} \Omega_{uu}^{-1}) b_1 \\ \vdots \\ \text{row}_n(\Delta_{vv}) \text{col}_n(\Omega_{vv}^{-1} \Omega_{vu} \Omega_{uu}^{-1}) b_n \end{bmatrix}}_{\mathcal{B}_{vu}}.$$

Combining the results above leads to:

$$G_T Z' \widehat{\Sigma}_u^{-1}(q) u - G_T Z' (I_T \otimes \widehat{\Omega}_{uu}^{-1} \widehat{\Omega}_{uv} \widehat{\Omega}_{vv}^{-1}) v \Rightarrow \int_0^1 J(r) \Omega_{uu}^{-1} dB_{u,v}(r) + \mathcal{B}^+,$$

where $\mathcal{B}^+ := \mathcal{B}_{\epsilon\eta} - \mathcal{B}_{vu}$. By construction, we have $G_T \widehat{\mathcal{B}}^+ \Rightarrow \mathcal{B}^+$. Altogether this implies the limiting distribution in the theorem. ■

Proof of Theorem 3.4 We first introduce the appropriate scaling into the test statistic \mathcal{W} given by

$$(\mathbf{R} \widehat{\beta}_{FGLS}^+ - \mathbf{r})' (\mathbf{R} G_T^{-1} \mathbf{R}') \left[(\mathbf{R} G_T^{-1} \mathbf{R}') \widehat{\Phi} (\mathbf{R} G_T^{-1} \mathbf{R}') \right]^{-1} (\mathbf{R} G_T^{-1} \mathbf{R}') (\mathbf{R} \widehat{\beta}_{FGLS}^+ - \mathbf{r}).$$

Since the matrices G_T^{-1} and $\mathbf{R}' \mathbf{R}$ commute and $\mathbf{R} \mathbf{R}' = \mathbf{I}_k$, we have $(\mathbf{R} G_T^{-1} \mathbf{R}') (\mathbf{R} \widehat{\beta}_{FGLS}^+ - \mathbf{r}) = \mathbf{R} G_T^{-1} (\widehat{\beta}_{FGLS}^+ - \beta)$ under the null hypothesis. Conditional on $\mathcal{F}_v = \sigma(B_v(r), 0 \leq r \leq 1)$, this quantity is asymptotically normally distributed by Theorem 3.3(c) with asymptotic covariance matrix

$$\mathbf{R} \left(\int_0^1 J(r) \Omega_{uu}^{-1} J(r)' dr \right)^{-1} \left(\int_0^1 J(r) \Omega_{uu}^{-1} \Omega_{u,v} \Omega_{uu}^{-1} J(r)' dr \right) \left(\int_0^1 J(r) \Omega_{uu}^{-1} J(r)' dr \right)^{-1} \mathbf{R}'.$$

The consistent estimation of all the quantities involved ensures that $(\mathbf{R} G_T^{-1} \mathbf{R}') \widehat{\Phi} (\mathbf{R} G_T^{-1} \mathbf{R}')$ has the same limit. Therefore, the Wald statistics is conditionally chi-square distributed with k degrees of freedom. Since this distribution does not depend on \mathcal{F}_v , we know the unconditional distribution of \mathcal{W} is also χ_k^2 . ■

Proof of Theorem 3.5 The results for K_{j,b_T}^{SOLS} and K_{j,b_T}^{SUR} follow from a straightforward multivariate extensions of the proof of Proposition 6 in Wagner and Hong (2016). For K_{j,b_T}^{BIAM} , we first define the population counterparts of $\varphi_{j,b_T}(\{\hat{u}_{FGLS}\})$ and $\widehat{\Sigma}_u^{-1}(q_T, b_T)$. That is, let $\varphi_{j,b_T} := [u_j', \sum_{s=j}^{j+1} u_s', \dots, \sum_{s=j}^{j+b_T-1} u_s']'$ and let $\Sigma_u^{-1}(q_T, b_T)$ denote the

subblock matrix of $\Sigma_u^{-1}(q_T)$ formed by taking the elements with row and column indices belonging to the set $\{n(T - b_T) + 1, n(T - b_T) + 2, \dots, nT\}$. By rearrangement,

$$\begin{aligned} K_{j,b_T}^{BIAM} &= b_T^{-2} \varphi'_{j,b_T} \Sigma_u^{-1}(q_T, b_T) \varphi_{j,b_T} \\ &\quad + b_T^{-2} \varphi'_{j,b_T} \left(\widehat{\Sigma_u^{-1}}(q_T, b_T) - \Sigma_u^{-1}(q_T, b_T) \right) \varphi_{j,b_T} + R(q_T, j, b_T), \end{aligned} \quad (3.62)$$

where the remainder term is bounded as

$$\begin{aligned} |R(q_T, j, b_T)| &\leq \left\| b_T^{-1} \left(\varphi_{j,b_T}(\{\hat{\mathbf{u}}_{FGLS}\}) - \varphi_{j,b_T} \right) \right\|^2 \left\| \widehat{\Sigma_u^{-1}}(q_T, b_T) \right\| \\ &\quad + 2 \left\| b_T^{-1} \left(\varphi_{j,b_T}(\{\hat{\mathbf{u}}_{FGLS}\}) - \varphi_{j,b_T} \right) \right\| \left\| \widehat{\Sigma_u^{-1}}(q_T, b_T) \right\| \left\| b_T^{-1} \varphi_{j,b_T} \right\|. \end{aligned}$$

Poincaré's separation theorem (e.g. page 347-348 of Abadir and Magnus (2005)) implies $\lambda_{\min}(\mathbf{A}) \leq \lambda_{\min}(\mathbf{B}) \leq \lambda_{\max}(\mathbf{B}) \leq \lambda_{\max}(\mathbf{A})$ when \mathbf{B} is a principal submatrix of \mathbf{A} . By this inequality and Theorem 3.2, we conclude that $\|\widehat{\Sigma_u^{-1}}(q_T, b_T) - \Sigma_u^{-1}(q_T, b_T)\| \leq \|\Sigma_u^{-1}(q_T) - \Sigma_u^{-1}(q_T)\| = o_p(1)$ and $\|\widehat{\Sigma_u^{-1}}(q_T, b_T)\| = O_p(1)$. Moreover,

$$\left\| b_T^{-1} (\widehat{\varphi}_{j,b_T} - \varphi_{j,b_T}) \right\| = \left(b_T^{-1} \sum_{t=j}^{j+b_T-1} \left\| b_T^{-1/2} \sum_{s=j}^t (\widehat{\mathbf{u}}_{s,FGLS} - \mathbf{u}_s) \right\|^2 \right)^{1/2} = o_p(1)$$

where

$$\left\| b_T^{-1/2} \sum_{s=j}^t (\widehat{\mathbf{u}}_{s,FGLS} - \mathbf{u}_s) \right\| \leq \left\| b_T^{-1/2} \sum_{s=j}^t \mathbf{Z}'_s \mathbf{G}_T \right\| \left\| \mathbf{G}_T^{-1} (\widehat{\beta}_{FGLS}^+ - \beta) \right\| = o_p(1)$$

by Theorem 3.3 and the assumption $b_T/T \rightarrow 0$ as $T \rightarrow \infty$. Standard weak convergence arguments imply $\|b_T^{-1} \varphi_{j,b_T}\| = O_p(1)$. Combining these results, we have $b_T^{-2} \varphi'_{j,b_T} (\widehat{\Sigma_u^{-1}}(q_T, b_T) - \Sigma_u^{-1}(q_T, b_T)) \varphi_{j,b_T} = o_p(1)$ and $R(q_T, j, b_T) = o_p(1)$. By (3.62), it remains to consider $b_T^{-2} \varphi'_{j,b_T} \Sigma_u^{-1}(q_T, b_T) \varphi_{j,b_T}$. Construct the $nT \times nb_T$ selection matrix \mathbf{R}_{j,b_T} such that

$$\mathbf{R}_{j,b_T} \varphi_{j,b_T} = \left[\mathbf{0}', \dots, \mathbf{0}', \mathbf{u}'_j, \sum_{s=j}^{j+1} \mathbf{u}'_s, \dots, \sum_{s=j}^{j+b_T-1} \mathbf{u}'_s, \mathbf{0}', \dots, \mathbf{0}' \right]'$$

Then, by the MCD (3.7), we have

$$\begin{aligned} b_T^{-2} \varphi'_{j,b_T} \Sigma_u^{-1}(q_T, b_T) \varphi_{j,b_T} &= b_T^{-2} \varphi'_{j,b_T} \mathbf{R}'_{j,b_T} \Sigma_u^{-1}(q_T) \mathbf{R}_{j,b_T} \varphi_{j,b_T} \\ &= b_T^{-2} (\mathbf{M}_u(q_T) \mathbf{R}_{j,b_T} \varphi_{j,b_T})' \mathbf{S}_u^{-1}(q_T) (\mathbf{M}_u(q_T) \mathbf{R}_{j,b_T} \varphi_{j,b_T}). \end{aligned} \quad (3.63)$$

As argued in the proof of Theorem 3.1, by the assumption $\frac{q_T}{b_T} \rightarrow 0$ as $T \rightarrow \infty$, we can treat the premultiplication of $\mathbf{R}_{j,b_T} \boldsymbol{\varphi}_{j,b_T}$ by $\mathbf{M}_u(q_T)$ as applying the filter $\mathcal{A}_{q_T}(L)$ block-wise. Under the same condition, $\mathbf{S}_u^{-1}(q_T)$ implies a scaling $\mathbf{S}^{-1}(q_T)$. By the BN decomposition in Lemma 3.5, and similarly $\mathbf{C}(L) = \mathbf{C}(1) + (1 - L)\tilde{\mathbf{C}}(L)$ with $\mathbf{C}(L) = [\mathcal{A}(L)]^{-1}$,

$$\begin{aligned} \mathcal{A}_q(L) \sum_{s=j}^t \mathbf{u}_s &= \mathcal{A}_{q_T}(1) \sum_{s=j}^t \mathbf{u}_s + \tilde{\mathcal{A}}_{q_T}(L) \mathbf{u}_t \\ &= \mathcal{A}_{q_T}(1) \mathbf{C}(1) \sum_{s=j}^t \boldsymbol{\eta}_s + \tilde{\mathcal{A}}_{q_T}(1) \tilde{\mathbf{C}}(L) (\boldsymbol{\eta}_t - \boldsymbol{\eta}_{j-1}) + \tilde{\mathcal{A}}_{q_T}(L) \mathbf{u}_t. \end{aligned}$$

For $t = j + [rb_T] - 1$, a FCLT for i.i.d. sequences gives

$$b_T^{-1/2} \mathbf{S}^{-1/2}(q_T) \mathcal{A}_{q_T}(L) \sum_{s=j}^t \mathbf{u}_s = \mathbf{S}^{-1/2}(q_T) \mathcal{A}_{q_T}(1) \mathbf{C}(1) b_T^{-1/2} \sum_{s=j}^t \boldsymbol{\eta}_s + o_p(b_T^{-1/2}) \Rightarrow \mathbf{W}(r).$$

The partial sum process $\sum_{s=j}^t \boldsymbol{\eta}_s$ thus dominates the asymptotic distribution:

$$\begin{aligned} b_T^{-2} \boldsymbol{\varphi}_{j,b_T}' \boldsymbol{\Sigma}_u^{-1}(q_T, b_T) \boldsymbol{\varphi}_{j,b_T} \\ = b_T^{-1} \sum_{t=j}^{j+b_T-1} \left\| b_T^{-1/2} \mathbf{S}^{-1/2}(q_T) \mathcal{A}_{q_T}(L) \sum_{s=j}^t \mathbf{u}_s \right\|^2 + o_p(1) \Rightarrow \int_0^1 \|\mathbf{W}(r)\|^2 dr. \end{aligned}$$

An application of the continuous mapping theorem completes the proof. ■

Appendix 3.C Multivariate KPSS-type Tests

In the next theorem, we derive the analytical expression of the limiting distribution of multivariate KPSS-type tests.

Theorem 3.6

Let $\mathbf{W}(r) = [W_1(r), W_2(r), \dots, W_n(r)]'$ denote an n -dimensional standard Brownian motion. The cumulative density function (CDF) of $\int_0^1 \|\mathbf{W}(r)\|^2 dr$ is given by

$$F_n(x) = 2^{n/2} \sum_{j=0}^{\infty} k_{j,n} \operatorname{Erfc} \left(\frac{l_{j,n}}{2\sqrt{x}} \right), \quad x > 0,$$

where $k_{j,n} = (-1)^j \frac{\Gamma(j+n/2)}{j! \Gamma(n/2)}$, $l_{j,n} = 2\sqrt{2}j + \frac{n}{\sqrt{2}}$ and $\text{Erfc}(x) = \frac{2}{\sqrt{\pi}} \int_x^\infty e^{-t^2} dt$.

Proof We follow the approach from Example 1 of Anderson and Darling (1952). Let f_n denote the probability density function of $\int_0^1 \|\mathbf{W}(r)\|^2 dr$ and write $\mathcal{L}\{\cdot\}$ and $\mathcal{L}^{-1}\{\cdot\}$ for the Laplace and inverse Laplace operator, respectively. By the equality $\int_0^1 \|\mathbf{W}(r)\|^2 dr = \sum_{i=1}^n \int_0^1 W_i(r)^2 dr$, independence of the components of $\mathbf{W}(r)$, and the known univariate result in Choi and Saikkonen (2010), we have

$$\mathcal{L}\{f_n(x)\}(t) = \int_0^\infty e^{-tx} f_n(x) dx = [\cosh(\sqrt{2t})]^{-n/2}. \quad (3.64)$$

According to equation (4.28) in Anderson and Darling (1952), the CDF is

$$\begin{aligned} F_n(x) &= \mathcal{L}^{-1} \left\{ \frac{1}{t} [\cosh(\sqrt{2t})]^{-n/2} \right\} (x) \\ &= \mathcal{L}^{-1} \left\{ \frac{1}{t} \left(\frac{e^{\sqrt{2t}}}{2} \right)^{-n/2} [1 + e^{-2\sqrt{2t}}]^{-n/2} \right\} (x) \\ &= \mathcal{L}^{-1} \left\{ \frac{1}{t} \left(\frac{e^{\sqrt{2t}}}{2} \right)^{-n/2} \sum_{j=0}^\infty k_{j,n} e^{-2j\sqrt{2t}} \right\} (x) \\ &= 2^{n/2} \sum_{j=0}^\infty k_{j,n} \mathcal{L}^{-1} \left\{ \frac{1}{t} e^{-l_{j,n}\sqrt{t}} \right\} (x) \end{aligned}$$

where we use (i) a t with a positive real part, (ii) linearity of the inverse Laplace operator, and (iii) the binomial expansion of $[1 + e^{-2\sqrt{2t}}]^{-n/2}$. The identity from Choi and Saikkonen (2010), $\mathcal{L}^{-1} \left\{ \frac{1}{t} e^{-u\sqrt{t}} \right\} (x) = 1 - \text{Erf} \left(\frac{u}{2\sqrt{x}} \right) = \text{Erfc} \left(\frac{u}{2\sqrt{x}} \right)$. ■

Appendix 3.D Estimating Quantities for FM Inference

The FM-GLS estimator relies on $\mathbf{\Omega}$, $\mathbf{\Delta}$, and $\mathbb{E}(\zeta_t \zeta_t')$ (see Assumption 3.1). For convenience, we denote this $(2n \times 2n)$ matrix $\begin{bmatrix} \Sigma_{\eta\eta} & \Sigma_{\eta\epsilon} \\ \Sigma_{\epsilon\eta} & \Sigma_{\epsilon\epsilon} \end{bmatrix}$ by Σ . Please note the difference between Σ and the large-dimensional matrix Σ_u . In this section, we consider the estimation of these three quantities within the BIAM framework. For convenience, we recall $\xi_t = [u_t', v_t']'$ and define $\xi = [\xi_1', \xi_2', \dots, \xi_T']'$. Similarly to the definition of Σ_u , we used $\Sigma_\xi := \mathbb{E}(\xi \xi')$ to denote the $(2nT \times 2nT)$

autocovariance matrix of $\{\xi_t\}$. As a sample counterpart, we stack $\widehat{\mathbf{u}}_t$ and $\Delta \mathbf{x}_t = \mathbf{v}_t$ in the $2n$ -dimensional vector $\widehat{\xi}_t = [\widehat{\mathbf{u}}_t', \mathbf{v}_t']'$. Using $\{\widehat{\xi}_t\}_{t=1}^T$, the BIAM estimator for Σ_ξ is now constructed as $\widehat{\Sigma}_\xi(q) = \widehat{\mathbf{M}}_\xi^{-1}(q) \widehat{\mathbf{S}}_\xi(q) \widehat{\mathbf{M}}_\xi^{-1'}(q)$, where the matrices $\widehat{\mathbf{M}}_\xi(q)$ and $\widehat{\mathbf{S}}_\xi(q)$ are defined similarly to $\widehat{\mathbf{M}}_u(q)$ and $\widehat{\mathbf{S}}_u(q)$, respectively. Since the BIAM estimator is fitting VAR processes up to order q_T (see (3.15)), the coefficient estimates $\widehat{\mathbf{F}}_j(q_T)$ of the j^{th} lag when a $\text{VAR}(q_T)$ is fitted, $j = 1, 2, \dots, q_T$, are immediate byproducts of the BIAM procedure and can thus be used to construct our estimators. Finally, if $\mathcal{F}(L) = \text{diag}[\mathcal{A}(L), \mathcal{D}(L)] := \mathbf{I}_{2n} - \sum_{j=1}^{\infty} \mathbf{F}_j L^j$, where $\mathbf{F}_j = \text{diag}[\mathbf{A}_j, \mathbf{D}_j]$, then $\mathcal{F}(L)\xi_t = \zeta_t$ holds.

Theorem 3.7

Recall the definitions $\widehat{\Omega}_{q_T} = \left(\mathbf{I}_{2n} - \sum_{j=1}^{q_T} \widehat{\mathbf{F}}_j(q_T)\right)^{-1} \widehat{\Sigma}_{q_T} \left(\mathbf{I}_{2n} - \sum_{j=1}^{q_T} \widehat{\mathbf{F}}_j(q_T)\right)^{-1'}$, $\widehat{\Delta}_{q_T, r_T} = \mathbf{Q}'_{r_T} \widehat{\Sigma}_\xi(q_T) \mathbf{Q}_1$, and

$$\begin{aligned} \widehat{\Sigma}_{q_T} = \frac{1}{T - q_T} \sum_{t=q_T+1}^T & \left[\widehat{\xi}_t - \widehat{\mathbf{F}}_1(q_T) \widehat{\xi}_{t-1} - \dots - \widehat{\mathbf{F}}_{q_T}(q_T) \widehat{\xi}_{t-q_T} \right] \\ & \times \left[\widehat{\xi}_t - \widehat{\mathbf{F}}_1(q_T) \widehat{\xi}_{t-1} - \dots - \widehat{\mathbf{F}}_{q_T}(q_T) \widehat{\xi}_{t-q_T} \right]', \end{aligned}$$

where $\mathbf{Q}_r = [\mathbf{O}_{2n \times 2n}, \dots, \mathbf{O}_{2n \times 2n}, \mathbf{I}_{2n}, \dots, \mathbf{I}_{2n}]'$ is an $(2nT \times 2n)$ block matrices of zeros of which the last r blocks have been replaced by identity matrices. If Assumptions 3.1-3.3 and 3.5 hold, then

$$\left\| \widehat{\Sigma}_{q_T} - \Sigma \right\| = O_p\left(\frac{q_T}{\sqrt{T}}\right) + O\left(\sum_{s=q_T+1}^{\infty} \|\mathbf{F}_s\|_{\mathcal{F}}^2\right) = o_p(1), \quad (3.65)$$

$$\left\| \widehat{\Omega}_{q_T} - \Omega \right\| = O_p\left(\sqrt{\frac{q_T^3}{T}} + \frac{1}{q_T} \sum_{s=q_T+1}^{\infty} s \|\mathbf{F}_s\|_{\mathcal{F}}\right) = o_p(1), \quad (3.66)$$

$$\left\| \widehat{\Delta}_{q_T, r_T} - \Delta \right\| = o(r_T^{-1}) + O_p\left(\sqrt{\frac{q_T^3}{T}} r_T + \sqrt{\frac{r_T}{q_T}} \sum_{s=q_T+1}^{\infty} s \|\mathbf{F}_s\|_{\mathcal{F}}\right) = o_p(1). \quad (3.67)$$

Proof Note that $\widehat{\xi}_t - \xi_t = [(\widehat{\mathbf{u}}_t - \mathbf{u}_t)', \mathbf{0}']'$ and hence $\|\widehat{\xi} - \xi\|^2 = \|\widehat{\mathbf{u}} - \mathbf{u}\|^2 = O_p(1)$ by Assumption 3.2. The conditions for Lemmas 3.3 – 3.4 and Theorem 3.2 are thus

satisfied and we can use these results in subsequent proofs. **(a)** The result (3.65) follows from the triangle inequality, Lemma 3.3 and (3.48). **(b)** The second result (3.66) is obtained by the definition $\Omega = (I_{2n} - \sum_{j=1}^{\infty} F_j)^{-1} \Sigma (I_{2n} - \sum_{j=1}^{\infty} F_j)^{-1'}$, Lemma 3.3 and a straightforward modification of (3.61). **(c)** By $\Delta = \sum_{h=r_T}^{\infty} \mathbb{E}(\xi_t \xi'_{t+h}) + Q'_{r_T} \Sigma_{\xi} Q_1$, the LHS of (3.67) can be bounded

$$\|\widehat{\Delta}_{q_T, r_T} - \Delta\| \leq \sum_{h=r_T}^{\infty} \|\mathbb{E}(\xi_t \xi'_{t+h})\|_{\mathcal{F}} + \|Q'_{r_T} \Sigma_{\xi}\| \|\widehat{\Sigma}_{\xi}^{-1}(q_T) - \Sigma_{\xi}^{-1}\| \|\widehat{\Sigma}_{\xi}(q_T) Q_1\|.$$

Since summability conditions on the coefficient matrices carry over to the autocovariances, we have $\sum_{h=r_T}^{\infty} \|\mathbb{E}(\xi_t \xi'_{t+h})\|_{\mathcal{F}} \leq r_T^{-1} \sum_{h=r_T}^{\infty} h \|\mathbb{E}(\xi_t \xi'_{t+h})\|_{\mathcal{F}} = o(r_T^{-1})$ by Assumption 3.1. Moreover, $\|Q'_{r_T} \Sigma_{\xi}\| \leq C \sqrt{r_T}$ and $\|\widehat{\Sigma}_{\xi}^{-1}(q_T) - \Sigma_{\xi}^{-1}\|$ is discussed in Theorem 3.2. Finally, showing $\|\widehat{\Sigma}_{\xi}(q_T) Q_1\| = O_p(1)$ will complete the proof after a straightforward comparison of the established stochastic orders. It suffices to prove $\|\widehat{\Sigma}_{\xi}(q_T)\| = O_p(1)$. Weyl's inequality (e.g. pages 40 and 46 in Tao (2012)) and Theorem 3.2 imply

$$\left| \lambda_{\min}(\widehat{\Sigma}_{\xi}^{-1}(q_T)) - \lambda_{\min}(\Sigma_{\xi}^{-1}) \right| \leq \|\widehat{\Sigma}_{\xi}^{-1}(q_T) - \Sigma_{\xi}^{-1}\| = o_p(1).$$

By the uniform boundedness of $\|\Sigma_{\xi}\|$, for a sufficiently large T , there exists a constant $C > 0$ such that $\|\widehat{\Sigma}_{\xi}(q_T)\|^{-1} = \lambda_{\min}(\widehat{\Sigma}_{\xi}^{-1}(q_T)) \leq C$ and thus $\|\widehat{\Sigma}_{\xi}(q_T)\| \leq C^{-1}$ with arbitrarily high probability. ■

Appendix 3.E Details on Implementation

The implementation of the BIAM estimator and the subsampling KPSS tests requires selecting the banding parameter q and the block length b . In our simulations and empirical application, we follow the subsampling and risk-minimization approach previously used by Bickel and Levina (2008), Wu and Pourahmadi (2009) and Ing et al. (2016) to select q . The steps are as follows:

Step 1 Split the series of (first-step OLS) residuals, $\{\widehat{\mathbf{u}}_t\}_{t=1}^T$, into J_0 non-overlapping subsequences of length l_0 . These subsequences are $\{\widehat{\mathbf{u}}_t\}_{t=(j-1)l_0+1}^{jl_0}$ for $j = 1, \dots, J_0$ with $J_0 = \lceil T/l_0 \rceil$.

Step 2 Select an integer H , $1 \leq H < l_0$, and construct the $(nH \times nH)$ sample autocovariance matrix $\widehat{\Pi}_{u,nH} = \frac{1}{T-H} \sum_{t=H}^{T-1} \widehat{u}_t(H) \widehat{u}_t(H)'$ which is an estimator of $\Sigma_{u,nH} := \mathbb{E}[u(H)u(H)']$ with $u(H) := \begin{bmatrix} u_1 \\ \vdots \\ u_H \end{bmatrix}$, where $\widehat{u}_t(\ell) = (\widehat{u}'_t, \dots, \widehat{u}'_{t-\ell+1})'$. Compute $\widehat{\Pi}_{u,nH}^{-1}$.

Step 3 For every subsequence of residuals $1 \leq j \leq J_0$, compute the BIAM estimate of $\Sigma_{u,nH}$ repeatedly for all possible banding parameters $1 \leq \bar{q} < H$, denoted as $\widehat{\Sigma}_{u,nH}^{-1}(\bar{q}; j)$.

Step 4 Select the banding parameter that minimizes the feasible average risk

$$q := \arg \min_{\bar{q} \in [1, H)} \frac{1}{J_0} \sum_{j=1}^{J_0} \left\| \widehat{\Sigma}_{u,nH}^{-1}(\bar{q}; j) - \widehat{\Pi}_{u,nH}^{-1} \right\|_p.$$

We take $p = 1$, $H = \lceil 2 \times T^{1/4} \rceil$ and $l_0 = \lceil T/5 \rceil$ and obtain satisfactory results for all the settings we have explored. As mentioned in Bickel and Levina (2008), the use of another vector norm (e.g. $p = 2$) does not lead to qualitatively different results. When we implement the minimum volatility rule as mentioned in Section 3.3.4 to select b , the values of tuning parameters are adopted from Wagner and Hong (2016), see their online supplement.

Appendix 3.F Additional Information for EKC

Appendix 3.F.1 Model fit

In Figure 3.5 below, we plot the fit of CO₂ emissions using the FM-SOLS, FM-SUR, and FM-GLS estimates.

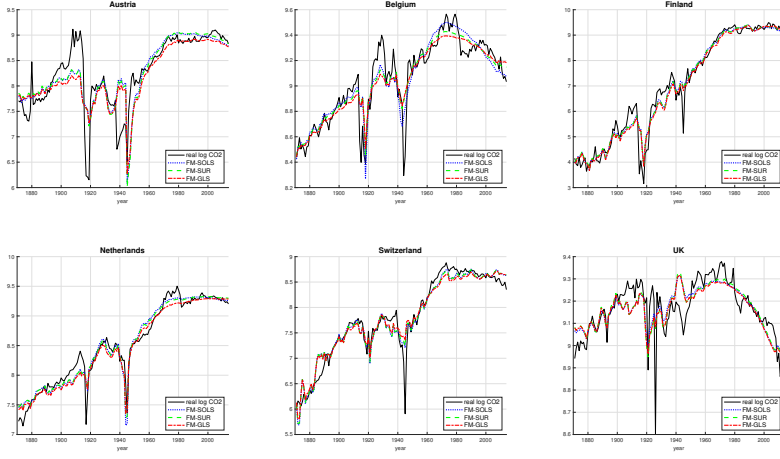


Figure 3.5: The fit of the FM-SOLS, FM-SUR, and FM-GLS estimates.

Appendix 3.F.2 Simulation DGP

The following procedure was used to obtain a simulation DGP that closely mimics the data characteristics.

- (a) Fit $\text{VAR}(p)$ models ($1 \leq p \leq 8$) to the series $\{\hat{u}_{t,FGLS}\}$ and $\{\Delta x_t\}$ individually. The BIC criterion select the $\text{VAR}(1)$ specification for both series (Table 3.7). Store the coefficient matrices \hat{A}_u and \hat{A}_v as well as the residual series $\{\hat{\eta}_t\}$ and $\{\hat{\varepsilon}_t\}$, respectively.

- (b) Stack $\hat{\zeta}_t = [\hat{\eta}_t', \hat{\varepsilon}_t']'$ and compute $\hat{\Sigma} = \frac{1}{T} \sum_{t=2}^T \hat{\zeta}_t \hat{\zeta}_t'$.

- (c) Denoting the estimated coefficients from the data by $\hat{\beta}_{FGLS}^+$, we generate the new data according to the following equations:

$$y_t = Z_t' \hat{\beta}_{FGLS}^+ + u_t,$$

with $\Delta \mathbf{x}_t = \mathbf{v}_t$, $\mathbf{x}_0 = \mathbf{0}$, and conditional on the estimates above,

$$\boldsymbol{\xi}_t = \begin{bmatrix} \mathbf{u}_t \\ \mathbf{v}_t \end{bmatrix} = \begin{bmatrix} \widehat{\mathbf{A}}_u & \mathbf{0} \\ \mathbf{0} & \widehat{\mathbf{A}}_v \end{bmatrix} \boldsymbol{\xi}_{t-1} + \boldsymbol{\zeta}_t, \text{ with } \boldsymbol{\zeta}_t \stackrel{i.i.d.}{\sim} \mathbf{N}(\mathbf{0}, \widehat{\boldsymbol{\Sigma}}).$$

Table 3.7: The numerical values for the BIC criterion.

	$p = 1$	$p = 2$	$p = 3$	$p = 4$	$p = 5$	$p = 6$	$p = 7$	$p = 8$
$\hat{\mathbf{u}}_{t,FGLS}$	-23.88	-23.49	-22.75	-21.87	-21.25	-20.66	-20.11	-19.27
$\Delta \mathbf{x}_t$	-37.94	-37.78	-37.15	-36.44	-35.68	-35.03	-34.34	-33.64

4

Cointegrating Polynomial Regressions with Power Law Trends: Environmental Kuznets Curve or Omitted Time Effects?

This chapter is based on the paper Lin and Reuvers (2020a).

Abstract

The Environment Kuznets Curve (EKC) predicts an inverted U-shaped relationship between economic growth and environmental pollution. Current analyses frequently employ models which restrict nonlinearities in the data to be explained by the economic growth variable only. We propose a Generalized Cointegrating Polynomial Regression (GCPR) with flexible time trends to provide an alternative explanation for the nonlinearities in the data. More specifically, a GCPR includes flexible powers of deterministic trends and integer powers of stochastic trends. We estimate this GCPR by nonlinear least squares and derive its asymptotic distribution. Endogeneity of the regressors can introduce nuisance parameters into this limiting distribution but a simulated approach nevertheless enables us to conduct valid inference. Moreover, a subsampling KPSS test is proposed to verify the validity of the cointegrating relation. The simulation study shows good performance of the simulated inference approach and subsampling KPSS test. We illustrate the GCPR approach using a dataset on GDP and CO₂ emissions for 18 industrialised countries. For each of these countries, a *linear* cointegrating relation between GDP and CO₂ around a power law trend is shown to provide an accurate description of the data. This suggests that the environmental improvement of the last years are due to economic factors different from GDP.

4.1 Introduction

On page 370 of their seminal paper, Grossman and Krueger (1995) conclude:

“Contrary to the alarmist cries of some environmental groups, we find no evidence that economic growth does unavoidable harm to the natural habitat. Instead we find that while increases in GDP may be associated with worsening environmental conditions in very poor countries, air and water quality appear to benefit from economic growth once some critical level of income has been reached.”

The quote above suggests an inverted U-shaped relationship between environmental degradation and economic growth. This relationship is currently known as the Environmental Kuznets Curve (EKC) and it forms an active research area. Its relevance becomes clear if we look at some forecasts of long-run economic growth. The projected GDP per capita growth of the world is about 2.1% per year for the next decades (see, e.g., chapter 3 in Nordhaus (2013) or Gillingham and Nordhaus (2018)) and this growth is (partially) powered by carbon-based energy resources, water usage, and material consumption. In absence of an EKC, economic growth will place more and more stress on the environment. Alternatively, if the EKC exists, then the inverted U-shape eventually implies a turning point after which economic growth and environmental improvement go hand in hand. Due to such considerations, there is now, some 25 years after its first conception, a rich literature that (1) reports on the experimental evidence on the existence/nonexistence of the EKC, (2) provides economic theory to explain the EKC, and/or (3) refines the econometric tools that are used to analyse the EKC.¹ To quantify the volume of the literature we have entered the search query “*Environmental Kuznets Curve*” into the Web of Science: more than 3,300 references are found.²

¹Further references to these specific areas of research can be found in the review articles by Dasgupta et al. (2002), Stern (2004), and Carson (2009) among others.

²Web of Science, accessed on January 29, 2021, <http://www.webofknowledge.com>.

The studies on the EKC have been criticised on two main points. First, the GDP variable was initially treated as a stationary variable even though unit root tests do not reject the null hypothesis. This has further consequences since EKC regressions include higher integer powers of GDP as well. The combination of nonstationarity and nonlinearity places the EKC in the nonlinear cointegration literature and appropriate econometric techniques should be employed. Such techniques have been provided in Wagner (2015) and Wagner and Hong (2016) under the name of *Cointegrating Polynomial Regressions* (CPRs), that is regressions containing: deterministic variables, integrated processes, and integer powers of integrated processes.³ These CPRs are estimated by fully modified OLS to allow for standard inference on the coefficients. Seemingly unrelated regressions to analyse multiple countries are discussed in Wagner et al. (2020) and Chapter 3.

As a second point of critique, there is an ongoing debate on the model specification. First, one can consider various functional forms when modeling the relationship between the GDP and the pollution variable. A quadratic specification is widely used but a cubic relation (Wagner (2015)) and double-nonlinear transformations (Lin et al. (2020)) have also been proposed. Various specification tests are available to decide on the right specification (see, e.g. Hong and Phillips (2010), Wang and Phillips (2012) and Wang et al. (2018)), or one could resort to nonparametric estimation procedures as in Wang and Phillips (2009) or Linton and Wang (2016). Whereas such modelling approaches do allow for a more flexible relationship between GDP and the pollution index, they also implicitly assume that nonlinear environmental effects are explainable by economic growth only. However, relevant variables are possibly missing from the model specification. Such omitted variables are a valid concern because

³The recent work by Stypka et al. (2017) confirms the importance of treating the growth variable as nonstationary. However, it appears less important to use an estimation procedure that incorporates the fact that several integer powers of the same integrated process appear as regressors. Namely, Stypka et al. (2017) also find that the “standard estimator” which treats higher order powers of the integrated regressor as additional $I(1)$ variables has the same limiting distribution as the CPR estimator (yet a slightly worse finite sample performance).

adaptation to clean technology,⁴ pollution control policy,⁵ increasing energy efficiency, and increasing environmental awareness may all influence pollution levels yet are also difficult to quantify or available for short time spans only (and for that reason often excluded from the reduced-form model). It has been argued that the inclusion of deterministic time trends will control for such omitted variables. In empirical applications, this typically translates into the inclusion of a linear deterministic trend, see Panayotou (1997) and Stypka et al. (2017), for example.

There seems no a priori reason why *linear* deterministic trends should control for omitted variables and provide a valid EKC specification. On the contrary, we will reason here, and later also in the empirical application, that omitted nonlinear trends are more likely to result in erroneous EKC results. The small simulation setting in Table 4.1 illustrates this point. We consider an omitted nonlinear deterministic trend and estimate an EKC-type of regression: $y_t = \tau_1 + \tau_2 t + \phi_1 x_t + \phi_2 x_t^2 + u_t$, where x_t and y_t are variables measuring economic growth and environmental pollution, respectively. We test $H_0 : \phi_2 \geq 0$ versus $H_a : \phi_2 < 0$ because a significantly negative coefficient in front of x_t^2 is the typical result economists interpret as the existence of the EKC. It is seen how a (correctly sized) Wald test misinterprets the negative curvature of a deterministic trend for negative curvature caused by a squared integrated variable (Panel (A)). In other words, a negative and significant coefficient in front of the square of GDP might be caused by omitted nonlinear deterministic trends rather than being any evidence for the EKC. To be on the safe side we thus recommend researchers to include a nonlinear trend component in their model specification. This is rather innocuous. Indeed, Panel (B) of Table 4.1 shows that: (1) significant results for nonlinear economic growth effects continue to be found when ϕ is indeed unequal to zero, and (2) the added

⁴Nordhaus (2014) discusses the link between climate change and technological changes. As another example, Figure 2 in Gillingham and Stock (2018) reports a steady decline in the price of solar panels and a steady growth in solar panel sales. Cheaper solar energy can substitute fossil energy thereby reducing pollution.

⁵A policy variable, ‘Repudiation of Contracts by Government’, was included by Panayotou (1997) to proxy the quality of environmental policies and institutions.

nonlinear deterministic trend has only a minor influence on statistical power.

Next to the well-developed literature on the different ways of incorporating the GDP variable (see above), we will investigate the influence of more flexible deterministic trend specifications. As such, this chapter augments the Cointegrating Polynomial Regressions of Wagner and Hong (2016) with power law deterministic trends. These power law trends are used in the literature to allow for non-constant growth rates in technology indices (see e.g. Duggal et al. (1999) and Duggal et al. (2007)) or production functions (Klein et al. (2004)). Alternatively, as in Li and Linton (2021), the reader can also view these flexible trends as a purely statistical method to model the data rather than having any direct interpretation in terms of the dynamics of the process. These trends now merely approximate missing variables and provide an outside option (next to the GDP variable) to describe nonlinearities in the data. Theoretical properties of models with power law trends have been derived in Phillips (2007), Robinson (2012) and Gao et al. (2020). Building upon this strand of literature, we provide the limiting distribution of the estimator and propose a simulated approach for parameter inference. Additionally, we show how a KPSS-type of test remains useful in verifying the stationarity of the error process hence avoiding spurious results or misspecification of the cointegrating relation. A Monte Carlo study sheds light on the finite sample properties of the simulated approach and stationarity test. As an empirical application, we look at a dataset on 18 countries over the timespan 1870-2014 and study the Environmental Kuznets Curve. For all 18 countries, we find that the flexible deterministic trends sufficiently capture the nonlinearities in the data and turn higher integer powers of log per capita GDP redundant. We thus recommend researchers to check whether their EKC conclusions are robust to the inclusion of power law trends.

This chapter is organized as follows. Section 4.2 introduces the model and the estimation framework. Asymptotic properties of the estimators and parameter inference are discussed in Section 4.3. The Monte Carlo simulations in Section

Table 4.1: The rejection rate (in %) of a t-test for $H_0 : \phi_2 \geq 0$ versus $H_1 : \phi_2 < 0$. **(A)** A missing nonlinear deterministic trend in the model causing falsely inflated rejections of $H_0 : \phi_2 = 0$ versus $H_a : \phi_2 < 0$. **(B)** Adding an additional quadratic deterministic trend to the model specification hardly influences the power of the test $H_0 : \phi_2 = 0$ vs. $H_a : \phi_2 \neq 0$. That is, a significant coefficient in front of x_t^2 remains after adding a redundant quadratic trend.

Panel (A): Omitted t^2 in EKC Specification						
DGP: $y_t = 7 + 0.05t + \tau t^2 + 5x_t + u_t$						
Model 1: $y_t = \tau_1 + \tau_2 t + \phi_1 x_t + \phi_2 x_t^2 + u_t$						
τ	0	-0.0001	-0.0002	-0.0003	-0.0004	-0.0005
$T = 100$	5.58	8.50	13.42	19.41	27.06	32.89
$T = 200$	5.08	37.67	57.88	64.92	69.14	70.79
Panel (B): Redundant t^2 in EKC Specification						
DGP: $y_t = 7 + 0.05t + 5x_t + \phi x_t^2 + u_t$						
Model 2: $y_t = \tau_1 + \tau_2 t + \phi_1 x_t + \phi_2 x_t^2 + u_t$						
Model 3: $y_t = \tau_1 + \tau_2 t + \tau_3 t^2 + \phi_1 x_t + \phi_2 x_t^2 + u_t$						
ϕ	0	-0.1	-0.2	-0.3	-0.4	-0.5
$T = 100$	5.47	42.85	70.66	84.70	91.64	95.89
	(5.65)	(36.10)	(66.23)	(82.27)	(90.46)	(95.21)
$T = 200$	5.24	83.02	97.14	99.32	99.82	99.97
	(5.31)	(80.13)	(96.66)	(99.26)	(99.80)	(99.97)

Note: We have used $\begin{bmatrix} u_t \\ v_t \end{bmatrix} \stackrel{i.i.d.}{\sim} N(\mathbf{0}, \mathbf{I}_2)$ and $x_t = \sum_{s=1}^t v_s$. In panel (B), we have placed the results for Model 3 within parentheses.

4.4 compare asymptotic results and the finite sample distributions. An in depth discussion of the Environment Kuznets Curve can be found in Section 4.5. Section 4.6 concludes. All proofs can be found in the Appendix.

Finally, some words on notation. The integer part of the number $a \in \mathbb{R}^+$ is denoted by $[a]$. For a vector $x \in \mathbb{R}^n$, its p -norm is denoted by $\|x\|_p = (\sum_{i=1}^n |x_i|^p)^{1/p}$. For a matrix A , say of dimension $(n \times m)$, the induced p -norm and Frobenius norm are defined as $\|A\|_p = \sup_{x \neq 0} \|Ax\|_p / \|x\|_p$ and $\|A\|_F = \sqrt{\sum_{i=1}^n \sum_{j=1}^m a_{ij}^2}$, respectively. For p -norms, we will omit the subscripts whenever $p = 2$. I_n denotes the $(n \times n)$ identity matrix. Two special linear algebra operators are: the Hadamard product (element-wise multiplication) denoted by “ \odot ” and the Kronecker product denoted by “ \otimes ”. We omit the integration bounds whenever the integral is take over $[0, 1]$. The symbol “ \Rightarrow ” signifies weak convergence, “ $\stackrel{d}{=}$ ” stands for equality in distribution, and “ \rightarrow_p ” and “ \rightarrow_d ” denote convergence in probability and in distribution. If convergence occurs conditionally on the sample, then we add a superscript “*” to the standard notation. The probabilistic Landau symbols are $O_p(\cdot)$ and $o_p(\cdot)$. Finally, the generic constant C can change from line to line.

4.2 The Model and NLS Estimation

Our model specification combines power law regressions with the cointegrating polynomial regression (CPR) introduced by Wagner and Hong (2016). It combines integrated regressors (and their integer powers) with a flexible deterministic trend specification. The resulting *Generalized Cointegrating Polynomial Regression* (GCPR) for y_t is given by

$$\begin{aligned} y_t &= \sum_{i=1}^d \tau_i t^{\theta_i} + \sum_{i=1}^m \sum_{j=1}^{p_i} \phi_{ij} x_{it}^j + u_t \\ &= d_t(\theta)' \tau + \sum_{i=1}^m x'_{(i)t} \phi_i + u_t = d_t(\theta)' \tau + s'_t \phi + u_t, \quad (4.1) \end{aligned}$$

where $\boldsymbol{\theta} = [\theta_1, \theta_2, \dots, \theta_d]'$, $\boldsymbol{\tau} = [\tau_1, \tau_2, \dots, \tau_d]'$, $\boldsymbol{\phi}_i = [\phi_{i1}, \phi_{i2}, \dots, \phi_{i, p_i}]'$, $\mathbf{d}_t(\boldsymbol{\theta}) = [t^{\theta_1}, \dots, t^{\theta_d}]'$ and $\mathbf{x}_{(i)t} = [x_{it}, x_{it}^2, \dots, x_{it}^{p_i}]'$ collects the integer powers of the i^{th} integrated regressor ($i = 1, 2, \dots, m$). The final equality in (4.1) relies on the definitions $\mathbf{s}_t = [\mathbf{x}'_{(1)t}, \mathbf{x}'_{(2)t}, \dots, \mathbf{x}'_{(m)t}]'$ and $\boldsymbol{\phi} = [\boldsymbol{\phi}'_1, \boldsymbol{\phi}'_2, \dots, \boldsymbol{\phi}'_m]'$. This notation is introduced for theoretical generality and it allows for a fixed number of d flexible deterministic trends to be incorporated into the model. Two remarks are of particular relevance for the applied researcher. First, we envision model specifications where d is small such that the deterministic trends cannot fully represent the integrated regressors (see, e.g. Phillips (1998)). Second, some elements of the vector $\boldsymbol{\theta}$ can be fixed prior to the analysis. For example, setting $\theta_1 = 0$ and $\theta_2 = 1$ automatically includes an intercept and linear trend in the model. A typical model specification is $y_t = \tau_1 + \tau_2 t + \tau_3 t^\theta + \mathbf{s}'_t \boldsymbol{\phi} + u_t$. Our approach and proofs are easily adapted to the case in which some elements of $\boldsymbol{\theta}$ are prespecified. The error term u_t is stationary (see Assumption 4.2 for more details).

We consider nonlinear least squares (NLS) estimators of the unknown parameters in (4.1). As such, we define the objective function $Q_T(\boldsymbol{\theta}, \boldsymbol{\tau}, \boldsymbol{\phi}) = \frac{1}{2} \sum_{t=1}^T (y_t - \mathbf{d}_t(\boldsymbol{\theta})' \boldsymbol{\tau} - \mathbf{s}'_t \boldsymbol{\phi})^2$ and compute

$$\left(\widehat{\boldsymbol{\theta}}_T, \widehat{\boldsymbol{\tau}}_T, \widehat{\boldsymbol{\phi}}_T \right) = \arg \min_{(\boldsymbol{\theta}, \boldsymbol{\tau}, \boldsymbol{\phi}) \in \Theta \times \mathbb{R}^d \times \mathbb{R}^p} Q_T(\boldsymbol{\theta}, \boldsymbol{\tau}, \boldsymbol{\phi}), \quad (4.2)$$

where $p = \sum_{i=1}^m p_i$,

$$\Theta = \left\{ \theta_1, \dots, \theta_d : -\frac{1}{2} < \theta_L \leq \theta_1; \theta_j - \theta_{j-1} \geq \delta, j = 2, \dots, d; \theta_d \leq \theta_U < \infty \right\},$$

for some lower bound θ_L , some upper bound θ_U , and $\delta > 0$. Note that $\Theta \subset \mathbb{R}^p$ is compact.

The optimization problem in (4.2) is easy to solve. For any given $\boldsymbol{\theta} \in \Theta$, the minimizers for $\boldsymbol{\tau}$ and $\boldsymbol{\phi}$ can be found from the OLS regression

$$\begin{bmatrix} \boldsymbol{\tau}(\boldsymbol{\theta}) \\ \boldsymbol{\phi}(\boldsymbol{\theta}) \end{bmatrix} = \left(\sum_{t=1}^T \mathbf{z}_t(\boldsymbol{\theta}) \mathbf{z}_t(\boldsymbol{\theta})' \right)^{-1} \left(\sum_{t=1}^T \mathbf{z}_t(\boldsymbol{\theta}) y_t \right), \text{ with } \mathbf{z}_t(\boldsymbol{\theta}) = [\mathbf{d}_t(\boldsymbol{\theta})', \mathbf{s}'_t]'. \quad (4.3)$$

We can thus minimize the concentrated criterion function $\widetilde{Q}_T(\theta) = Q_T(\theta, \tau(\theta), \phi(\theta))$ to obtain $\widehat{\theta}_T$ and run a subsequent OLS regression to find $\widehat{\tau}_T$ and $\widehat{\phi}_T$.

Remark 4.1

The fixed powers of x_{it} allow us to test for their significance and thereby distinguish between nonlinearities caused by deterministic and/or stochastic trends. This is important for our empirical application on the Environmental Kuznets Curve, see Section 4.5. Hu et al. (2019) study a model with a flexible power of the integrated regressor. That is, these authors derive the limiting distribution of the NLS estimators for β and γ when $y_t = \beta|x_t|^\gamma + u_t$ with $\beta \neq 0$.

4.3 Asymptotic Theory

We subsequently study the asymptotic properties of the NLS estimators. To this end we first collect all the unknown parameters in the vector $\gamma = [\theta', \tau', \phi']'$. This vector is assumed to be an element of the parameter space $\Gamma = \Theta \times \mathbb{R}^{d+p}$. The true parameter vector is $\gamma_0 = [\theta'_0, \tau'_0, \phi'_0]'$.

Assumption 4.1

For all $1 \leq i \leq d$, we have $\tau_{0i} \neq 0$.

Assumption 4.2

Let $\zeta_t = [\eta'_t, \varepsilon'_t]'$ be a sequence of i.i.d. random vectors with $\mathbb{E}(\zeta_t) = \mathbf{0}$, $\Sigma = \mathbb{E}(\zeta_t \zeta'_t)$, and $\mathbb{E} \|\zeta_t\|^{2q} < \infty$ for some $q > 2$.

- (a) $u_t = \sum_{k=0}^{\infty} \psi_k \eta_{t-k}$ with $\sum_{k=1}^{\infty} k|\psi_k| < \infty$.
- (b) $x_t = \sum_{s=1}^t v_s$, where $v_t = \sum_{k=0}^{\infty} \Psi_k \varepsilon_{t-k}$ with $\sum_{k=0}^{\infty} \|\Psi_k\| < \infty$ and $\det\left(\sum_{k=0}^{\infty} \Psi_k\right) \neq 0$.

The first assumption is needed to avoid identification issues. That is, if $\tau_{0i} = 0$ for some $i \in \{1, \dots, d\}$, then the corresponding θ_{0i} is not identified and the

Davies problem arises when testing $H_0 : \tau_i = 0$ (see Davies (1977, 1987)). We will not consider such difficulties in the current paper and this is reflected in our model specification (4.1). That is, we consider *flexible* powers of the deterministic trends but *fixed* powers of the stochastic trends thus allowing us to test zero restrictions on (elements of) ϕ . This is of crucial importance in the EKC application to see whether nonlinear effects in the economic growth variable remain significant after nonlinear time trends have been added to the model. For different model settings Assumption 4.1 has been relaxed in the literature. Baek et al. (2015) and Cho and Phillips (2018) study the asymptotic behaviour of a quasi-likelihood ratio test when Assumption 4.1 is violated and the conditional mean of the data contains strictly stationary regressors and a flexible time trend. Alternatively, one can use drifting parameter sequences with different identification strengths as in Andrews and Cheng (2012).

Assumption 4.2 excludes cointegration among elements of \mathbf{x}_t and defines it as the partial sum of a short memory process. The latter implies that $\frac{1}{\sqrt{T}} \sum_{t=1}^{\lfloor rT \rfloor} \begin{bmatrix} u_t \\ v_t \end{bmatrix} \Rightarrow B(r) = \begin{bmatrix} B_u(r) \\ B_v(r) \end{bmatrix}$ where $B(r)$ denotes an $(m+1)$ -dimensional vector Brownian motion with covariance matrix $\Omega = \begin{bmatrix} \Omega_{uu} & \Omega_{uv} \\ \Omega_{vu} & \Omega_{vv} \end{bmatrix}$. The one-sided long-run covariance matrix $\Delta = \sum_{h=0}^{\infty} \mathbb{E} \left(\begin{bmatrix} u_t u_{t+h} & u_t v'_{t+h} \\ v_t u_{t+h} & v_t v'_{t+h} \end{bmatrix} \right) = \begin{bmatrix} \Delta_{uu} & \Delta_{uv} \\ \Delta_{vu} & \Delta_{vv} \end{bmatrix}$ is partitioned similarly. Subscripts are used to refer to specific elements. For example, B_{v_i} and $\Delta_{v_i u}$ denote the i^{th} elements of B_v and Δ_{vu} , respectively.

A concise exposition of our results asks for additional notation. An enumeration of various definitions is presented below.

- (1) Introduce scaling matrices: $D_{d,T}(\theta) = \text{diag}[T^{\theta_1}, T^{\theta_2}, \dots, T^{\theta_d}]$ for the time trends and their coefficients, and $D_{s,T} = \text{diag}[D_{(1),T}, \dots, D_{(m),T}]$ for the integer powers of $I(1)$ regressors, where $D_{(i),T} = \text{diag}[T^{1/2}, T, \dots, T^{p_i/2}]$. Moreover, we define two $(2d+p) \times (2d+p)$ nonrandom block matrices $L_{\tau_0,T}$ and $D_{\theta_0,T}$ such that $L_{\tau_0,T} = \begin{bmatrix} I_d & -\text{diag}[\tau_0] \ln T & \mathbf{0} \\ \mathbf{0} & I_d & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & I_p \end{bmatrix}$, $D_{\theta_0,T} = \sqrt{T} \begin{bmatrix} D_{d,T}(\theta_0) & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & D_{d,T}(\theta_0) & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & D_{s,T} \end{bmatrix}$, and $G_{\gamma_0,T} = D_{\theta_0,T} L'_{\tau_0,T} = \sqrt{T} \begin{bmatrix} D_{d,T}(\theta_0) & \mathbf{0} & \mathbf{0} \\ D_{d,T}(\theta_0) \text{diag}[\tau_0] \ln T & D_{d,T}(\theta_0) & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & D_{s,T} \end{bmatrix}$.

(2) Define (nonrandom and stochastic) vectors $\mathbf{d}(r; \boldsymbol{\theta}_0) = [r^{\theta_{10}}, r^{\theta_{20}}, \dots, r^{\theta_{d0}}]'$, $\mathbf{B}_{(i)}(r) = [\mathbf{B}_{v_i}(r), \mathbf{B}_{v_i}^2(r), \dots, \mathbf{B}_{v_i}^{p_i}(r)]'$ and their stacked random vector process $\mathbf{j}(r; \boldsymbol{\gamma}_0) = [(\boldsymbol{\tau}_0 \odot \mathbf{d}(r; \boldsymbol{\theta}_0))' \ln(r), \mathbf{d}(r; \boldsymbol{\theta}_0)', \mathbf{B}'_{(1)}(r), \dots, \mathbf{B}'_{(m)}(r)]'$.

(3) We define $\mathbf{b}_i = [1, 2 \int_0^1 \mathbf{B}_{v_i}(r) dr, \dots, p_i \int_0^1 \mathbf{B}_{v_i}^{p_i-1}(r) dr]'$ and $\mathcal{B}_{vu} = [\mathbf{0}'_{d \times 1}, \mathbf{0}'_{d \times 1}, \mathbf{b}'_1 \boldsymbol{\Delta}_{v_{1u}}, \dots, \mathbf{b}'_m \boldsymbol{\Delta}_{v_{mu}}]'$ for the second-order bias terms.

Theorem 4.1

Under Assumptions 4.1-4.2, as $T \rightarrow \infty$, we have

$$\mathbf{G}_{\boldsymbol{\gamma}_0, T}(\widehat{\boldsymbol{\gamma}}_T - \boldsymbol{\gamma}_0) \Rightarrow \left(\int_0^1 \mathbf{j}(r; \boldsymbol{\gamma}_0) \mathbf{j}(r; \boldsymbol{\gamma}_0)' dr \right)^{-1} \left(\int_0^1 \mathbf{j}(r; \boldsymbol{\gamma}_0) dB_u(r) + \mathcal{B}_{vu} \right).$$

The proof of Theorem 4.1 is closely related to the work by Chan and Wang (2015). These authors provide the asymptotic distribution of NLS estimators with nonstationary time series under a set of general conditions (see their theorem 3.1). We verify that these conditions are also fulfilled when the scaling matrix $\mathbf{G}_{\boldsymbol{\gamma}_0, T}$ is non-diagonal and depending on the true parameter vector $\boldsymbol{\gamma}_0$. The results in Chan and Wang (2015) and Wang et al. (2018) suggest that Assumption 4.2 can be replaced by a long memory specification for $\Delta \mathbf{x}_t$. However, long memory parameters will enter the limiting distribution and inference will be complicated further.

We now illustrate Theorem 4.1 with two examples. These examples highlight two mathematical features that may complicate parameter inference.

Example 4.1

We consider the model $y_t = \tau t^\theta + u_t$ where the innovations satisfy Assumption 4.2. The limiting distribution of the parameter estimators depends solely on the mean square Riemann-Stieltjes integrals $\int_0^1 \tau_0 r^{\theta_0} \ln(r) dB_u$ and $\int_0^1 r^{\theta_0} dB_u$

and is therefore normally distributed (e.g., section 2.3 in Tanaka (2017)):

$$\begin{aligned} & \begin{bmatrix} T^{\theta_0 + \frac{1}{2}} & 0 \\ T^{\theta_0 + \frac{1}{2}} \tau_0 \ln(T) & T^{\theta_0 + \frac{1}{2}} \end{bmatrix} \begin{bmatrix} \widehat{\theta}_T - \theta_0 \\ \widehat{\tau}_T - \tau_0 \end{bmatrix} \\ & \Rightarrow N \left(\mathbf{0}, \Omega_{uu}(2\theta_0 + 1)^3 \begin{bmatrix} 2\tau_0^2 & -\tau_0(2\theta_0 + 1) \\ -\tau_0(2\theta_0 + 1) & (2\theta_0 + 1)^2 \end{bmatrix}^{-1} \right). \quad (4.4) \end{aligned}$$

Note especially how the scaling matrix in the LHS of (4.4), $\begin{bmatrix} T^{\theta_0 + \frac{1}{2}} & 0 \\ T^{\theta_0 + \frac{1}{2}} \tau_0 \ln(T) & T^{\theta_0 + \frac{1}{2}} \end{bmatrix}$, depends on θ_0 and is non-diagonal. The dependence on θ_0 is unavoidable but asymptotic results for the case of a diagonal scaling matrix are obtainable at the expense of a singular joint distribution. The latter follows directly from (4.4). That is, noting that $\begin{bmatrix} T^{\theta_0 + \frac{1}{2}} & 0 \\ 0 & T^{\theta_0 + \frac{1}{2}} / \ln(T) \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ -\tau_0 & 1/\ln(T) \end{bmatrix} \begin{bmatrix} T^{\theta_0 + \frac{1}{2}} & 0 \\ T^{\theta_0 + \frac{1}{2}} \tau_0 \ln(T) & T^{\theta_0 + \frac{1}{2}} \end{bmatrix}$ and $\lim_{T \rightarrow \infty} \begin{bmatrix} 1 & 0 \\ -\tau_0 & 1/\ln(T) \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ -\tau_0 & 0 \end{bmatrix}$, the continuous mapping theorem implies

$$\begin{bmatrix} T^{\theta_0 + \frac{1}{2}} & 0 \\ 0 & T^{\theta_0 + \frac{1}{2}} / \ln(T) \end{bmatrix} \begin{bmatrix} \widehat{\theta}_T - \theta_0 \\ \widehat{\tau}_T - \tau_0 \end{bmatrix} \Rightarrow \begin{bmatrix} 1/\tau_0 \\ -1 \end{bmatrix} \times N(\mathbf{0}, \Omega_{uu}(2\theta_0 + 1)^3).$$

This limiting distribution coincides with the result in theorem 6.3 of Phillips (2007).

Example 4.2

If $y_t = \tau t^\theta + \phi x_t + u_t$, then the limiting distribution of the NLS estimator is:

$$\begin{aligned} & \begin{bmatrix} T^{\theta_0 + \frac{1}{2}} & & \\ T^{\theta_0 + \frac{1}{2}} \tau_0 \ln(T) & T^{\theta_0 + \frac{1}{2}} & \\ & T & \end{bmatrix} \begin{bmatrix} \widehat{\theta}_T - \theta_0 \\ \widehat{\tau}_T - \tau_0 \\ \widehat{\phi}_T - \phi_0 \end{bmatrix} \\ & \Rightarrow \begin{bmatrix} \int (\tau_0 r^{\theta_0} \ln(r))^2 dr & \int \tau_0 r^{2\theta_0} \ln(r) dr & \int \tau_0 r^{\theta_0} \ln(r) B_v dr \\ \int \tau_0 r^{2\theta_0} \ln(r) dr & \int r^{2\theta_0} dr & \int r^{\theta_0} B_v dr \\ \int \tau_0 r^{\theta_0} \ln(r) B_v dr & \int r^{\theta_0} B_v dr & \int B_v^2 dr \end{bmatrix}^{-1} \left(\begin{bmatrix} \int \tau_0 r^{\theta_0} \ln(r) dB_u \\ \int r^{\theta_0} dB_u \\ \int B_v dB_u \end{bmatrix} + \begin{bmatrix} 0 \\ 0 \\ \Delta_{vu} \end{bmatrix} \right). \end{aligned}$$

This limiting distribution exhibits second order bias when $\Delta_{vu} \neq 0$, or when B_u and B_v are correlated.

Two features of the limiting distribution of $\mathbf{G}_{\gamma_0, T}(\widehat{\gamma}_T - \gamma_0)$ deserve further comments. First, as emphasised in Example 4.1, the scaling matrix $\mathbf{G}_{\gamma_0, T}$ features two uncommon properties: (1) this matrix depends on the true parameter vectors $\boldsymbol{\tau}_0$ and $\boldsymbol{\theta}_0$, and (2) $\mathbf{G}_{\gamma_0, T}$ is not diagonal. These peculiarities are caused by the nonlinearity and nonstationarity of the model. More specifically, these features can be traced back to the presence of functions like $f(t; \tau, \theta) = \tau t^\theta$. Limiting distributions with a similar mathematical structure can be found in the structural breaks literature, cf. model setting II.b of Perron and Zhu (2005) and its detailed analysis in Beutner et al. (2021).

Second, the nonstationary regressor x_{it} enters the model (4.1) through a polynomial transformation of the form $g(x, \phi_i) = \phi_{i1}x + \phi_{i2}x^2 + \dots + \phi_{ip_i}x^{p_i}$ ($i = 1, 2, \dots, m$). In the terminology of Park and Phillips (2001) this part of the regression function is a linear combination of H_0 -regular functions. It is well-documented in the literature, e.g. Chang et al. (2001) and Chan and Wang (2015), that this leads to second order bias terms and hence nonstandard inference (except in the special case of strictly exogenous nonstationary regressors).

4.3.1 General Considerations

Let us assume for the moment that $\boldsymbol{\theta}_0$ is known. The resulting model, that is $y_t = \mathbf{d}_t(\boldsymbol{\theta}_0)' \boldsymbol{\tau} + \mathbf{s}_t' \boldsymbol{\phi} + u_t$, is now linear in the unknown parameters $[\boldsymbol{\tau}', \boldsymbol{\phi}']'$. The OLS estimator is $\begin{bmatrix} \widehat{\boldsymbol{\tau}}_T(\boldsymbol{\theta}_0) \\ \widehat{\boldsymbol{\phi}}_T(\boldsymbol{\theta}_0) \end{bmatrix} = \left(\sum_{t=1}^T \mathbf{z}_t(\boldsymbol{\theta}_0) \mathbf{z}_t(\boldsymbol{\theta}_0)' \right)^{-1} \left(\sum_{t=1}^T \mathbf{z}_t(\boldsymbol{\theta}_0) y_t \right)$ and parameter inference is relatively straightforward. For example, we can apply the Fully Modified (FM) corrections as in Wagner and Hong (2016) to obtain a zero-mean Gaussian mixture limiting distribution that allows for standard inference.⁶

⁶The deterministic component in the CPR model by Wagner and Hong (2016) is a linear combination of the elements in the vector $[1, t, t^2, \dots, t^q]'$. The FM corrections are thus immediate if $\boldsymbol{\theta}_0$ is known and takes values in the natural numbers. Based on Lemma 4.2 in the Appendix, it is also relatively straightforward to derive such corrections when $\boldsymbol{\theta}_0$ is known but not necessarily elements of the natural numbers.

Given that θ_0 is unknown in practice, it seems natural to first compute $\widehat{\theta}_T$ by minimisation of $\widetilde{Q}_T(\theta)$, and to subsequently compute

$$\begin{bmatrix} \widehat{\tau}_T \\ \widehat{\phi}_T \end{bmatrix} = \left(\sum_{t=1}^T z_t(\widehat{\theta}_T) z_t(\widehat{\theta}_T)' \right)^{-1} \sum_{t=1}^T z_t(\widehat{\theta}_T) y_t. \quad (4.5)$$

The latter estimator is linear in y_1, y_2, \dots, y_T and fully modified adjustments seem possible. However, there are two issues. First, this estimator does not allow us to conduct inference on θ . Second, it is not completely clear how the estimation error in $\widehat{\theta}_T$ influences the limiting results.⁷

There is also some good news. Namely, if the estimator in (4.5) is used to calculate residuals, then these residuals can be used to construct consistent kernel estimators for the long-run variance (LRV) matrices Δ and Ω . With $V_t(\gamma) = [y_t - d_t(\theta)' \tau - s_t' \phi, \Delta x_t']'$, these estimators are defined as

$$\begin{aligned} \widehat{\Delta}_T &= \frac{1}{T} \sum_{t=1}^T \sum_{s=1}^t k\left(\frac{|t-s|}{b_T}\right) V_t(\widehat{\gamma}_T) V_s(\widehat{\gamma}_T)', \\ \widehat{\Omega}_T &= \frac{1}{T} \sum_{t=1}^T \sum_{s=1}^T k\left(\frac{|t-s|}{b_T}\right) V_t(\widehat{\gamma}_T) V_s(\widehat{\gamma}_T)', \end{aligned} \quad (4.6)$$

for some kernel function $k(\cdot)$ and bandwidth parameter b_T . The first element in $V_t(\widehat{\gamma}_T)$ is indeed the residual $\hat{u}_t = y_t - d_t(\widehat{\theta}_T)' \widehat{\tau}_T - s_t' \widehat{\phi}_T$. The remaining elements are $\Delta x_t = v_t$.

Assumption 4.3

(a) $k(0) = 1$, $k(\cdot)$ is continuous at zero, and $\sup_{x \geq 0} |k(x)| < \infty$.

(b) $\int_0^\infty \bar{k}(x) dx < \infty$, where $\bar{k}(x) = \sup_{y \geq x} |k(y)|$.

⁷We have investigated the asymptotic behaviour of a Fully Modified version of the estimator in (4.5). Our efforts in bounding the estimation error of $\widehat{\theta}_T$ lead to a term in the covariance asymptotics that is $O_p(\ln T)$ instead of $o_p(1)$ (also when imposing $\theta_L = 0$). This fact and the additional Monte Carlo simulations reported in Figure 4.3 in appendix suggests that this Fully Modified estimator is not asymptotically valid.

(c) The bandwidth parameters $\{b_T : T \geq 1\}$ satisfies $\{b_T\} \subseteq (0, \infty)$ and $\lim_{T \rightarrow \infty} (b_T^{-1} + T^{-1/2} b_T \ln T) = 0$.

The conditions on the kernel function $k(\cdot)$, Assumptions 4.3(a)-(b), are identical to those in Jansson (2002). Jansson (2002) remarks that these assumptions “*would appear to be satisfied by any kernel in actual use*”. Commonly used kernel functions such as the Bartlett, Parzen, and Quadratic Spectral kernels indeed satisfy all these assumptions. Assumption 4.3(c) differs from the usual requirement, $\lim_{T \rightarrow \infty} (b_T^{-1} + T^{-1/2} b_T) = 0$, by a factor $\ln T$. The difference is due to the estimation error in $\widehat{\theta}_T$. This error causes the residuals $\{\hat{u}_t\}$ to be less close to the innovations $\{u_t\}$ and we balance this by including autocovariance matrices of higher lags at a lower rate.

Theorem 4.2

Under Assumptions 4.1-4.3, we have $\widehat{\Delta}_T \rightarrow_p \Delta$ and $\widehat{\Omega}_T \rightarrow_p \Omega$.

4.3.2 Simulated inference

The limiting distribution in Theorem 4.1 is nonpivotal and thus unsuited for inference. We will use a simulated approach to account for the nuisance parameters, i.e. τ_0 , θ_0 and the parameters describing the covariance structure of $B(r)$. The main idea is to replace nuisance parameters by consistent estimates and to rely on a Monte Carlo (MC) simulation to approximate the limiting distribution. The empirical quantiles of these MC draws allow us to test hypothesis and/or conduct inference. Clearly, this kind of approach will provide exact inference when the limiting distribution is invariant with respect to the nuisance parameters (e.g. Dufour and Khalaf (2002) and Dufour (2006)). In the absence of such invariance, results as those in Wang et al. (2018) and Bergamelli et al. (2019) show that the simulation approach can retain an asymptotic justification. The following algorithm is an adaptation of the simulated estimation by Wang et al. (2018). Among others, the current setting

has to control for more nuisance parameters because of the flexible trend specification.

STEP 1: Estimate $\widehat{\gamma}_T$ and use the residuals $\{\hat{u}_t\}$ to compute the estimators $\widehat{\Delta}_T$ and $\widehat{\Omega}_T$ from (4.6)

STEP 2: Repeat for $j = 1, \dots, J$,

- (a) Draw an $(m + 1)$ -dimensional sequence $\{e_n\}_{n=1}^N$ i.i.d. from $N(\mathbf{0}, \mathbf{I}_{m+1})$.
- (b) Compute $\begin{bmatrix} \widehat{\mu}_n \\ \widehat{v}_n \end{bmatrix} = \widehat{\Omega}_T^{1/2} e_n$ and construct the partial sum process $\{\widehat{\chi}_n = [\widehat{\chi}_{1n}, \dots, \widehat{\chi}_{mn}]'\}_{n=1}^N$ according to $\widehat{\chi}_n = \widehat{\chi}_{n-1} + \widehat{v}_n$ and $\widehat{\chi}_0 = \mathbf{0}$.
- (c) Set $\widehat{w}_n = [n, \widehat{\chi}_n]'$, and construct a simulated draw as:

$$\widehat{\mathcal{J}}_N^{(j)}(\widehat{\gamma}_T, \widehat{\Omega}_T, \widehat{\Delta}_{vu}^-) = \left\{ \mathbf{G}_{\widehat{\gamma}, N}'^{-1} \left[\sum_{n=1}^N \dot{f}(\widehat{w}_n, \widehat{\gamma}_T) \dot{f}(\widehat{w}_n, \widehat{\gamma}_T)' \right] \mathbf{G}_{\widehat{\gamma}, N}^{-1} \right\}^{-1} \\ \times \left\{ \mathbf{G}_{\widehat{\gamma}, N}'^{-1} \left[\sum_{n=1}^N \dot{f}(\widehat{w}_n, \widehat{\gamma}_T) \widehat{\mu}_n \right] + \widehat{\mathcal{B}}_{vu}^- \right\},$$

where $\dot{f}(\widehat{w}_n, \gamma) = [(\tau \odot d_n(\theta))' \ln n, d_n(\theta)', \widehat{s}_n']'$ with $\widehat{s}_n = [\widehat{\chi}_{(1)n}, \widehat{\chi}_{(2)n}, \dots, \widehat{\chi}_{(m)n}]'$ and $\widehat{\chi}_{(i)t} = [\widehat{\chi}_{in}, \widehat{\chi}_{in}^2, \dots, \widehat{\chi}_{in}^{p_i}]'$, $i = 1, 2, \dots, m$. Moreover, $\widehat{\Delta}_{vu}^-$ is a consistent estimator of the subblock of $\Delta^- = \begin{bmatrix} \Delta_{uu}^- & \Delta_{uv}^- \\ \Delta_{vu}^- & \Delta_{vv}^- \end{bmatrix} = \Sigma - \Delta'$, and

$$\widehat{\mathcal{B}}_{vu}^- = [\mathbf{0}_{d \times 1}', \mathbf{0}_{d \times 1}', \widehat{b}_1' \widehat{\Delta}_{v_1 u}^-, \dots, \widehat{b}_m' \widehat{\Delta}_{v_m u}^-]'$$

$$\text{with } \widehat{b}_i = \left[1, 2 \frac{1}{N} \sum_{n=1}^N \left(\frac{\widehat{\chi}_{in}}{\sqrt{N}} \right), \dots, p_i \frac{1}{N} \sum_{n=1}^N \left(\frac{\widehat{\chi}_{in}}{\sqrt{N}} \right)^{p_i-1} \right]'$$

STEP 3: Use the empirical quantiles of elements of $\{\widehat{\mathcal{J}}_N^{(1)}, \dots, \widehat{\mathcal{J}}_N^{(J)}\}$ to conduct inference.

Step 3 has been kept general for notational convenience. A more concrete example is as follows. To construct a two-sided equal-tailed confidence interval for θ_{01} , we calculate the $\frac{\alpha}{2}$ and $1 - \frac{\alpha}{2}$ empirical quantile of the first elements of $\{\widehat{\mathcal{J}}_N^{(1)}, \dots, \widehat{\mathcal{J}}_N^{(J)}\}$, say $c_{\alpha/2}$ and $c_{1-\alpha/2}$ respectively. The implied confidence interval is $[\hat{\theta}_1 - c_{1-\alpha/2}T^{-(\hat{\theta}_1 + \frac{1}{2})}, \hat{\theta}_1 - c_{\alpha/2}T^{-(\hat{\theta}_1 + \frac{1}{2})}]$.

Theorem 4.3

Suppose Assumptions 4.1-4.3 hold and let $N = \lfloor \kappa T^\alpha \rfloor$ for some $\kappa > 0$ and $0 < \alpha \leq \min\{1, 1 + 2\tilde{\theta}\}$ with $\tilde{\theta} \in (-\frac{1}{2}, \theta_L)$. Then, we have

$$\begin{aligned} & \left\{ \mathbf{G}'_{\widehat{\gamma}, N} \left[\sum_{n=1}^N \dot{f}(\widehat{w}_n, \widehat{\gamma}_T) \dot{f}(\widehat{w}_n, \widehat{\gamma}_T)' \right] \mathbf{G}_{\widehat{\gamma}, N}^{-1} \right\}^{-1} \left\{ \mathbf{G}'_{\widehat{\gamma}, N}^{-1} \left[\sum_{n=1}^N \dot{f}(\widehat{w}_n, \widehat{\gamma}_T) \widehat{\mu}_n \right] + \widehat{\mathcal{B}}_{vu}^- \right\} \\ & \longrightarrow_{d^*} \left(\int_0^1 \mathbf{j}(r; \gamma_0) \mathbf{j}(r; \gamma_0)' dr \right)^{-1} \left(\int_0^1 \mathbf{j}(r; \gamma_0) dB_u(r) + \mathcal{B}_{vu} \right), \end{aligned}$$

in probability.

Theorem 4.3 establishes the asymptotic validity of the simulation approach. That is, for a large enough J , the empirical quantiles of the simulated distribution will coincide with the asymptotic distribution. According to Theorem 4.3, the length of the simulated time series should grow more slowly as θ_L approaches $-\frac{1}{2}$. For the general case with θ_L arbitrary close to $-\frac{1}{2}$ this is problematic since the allowed range for α will become arbitrary small. However, in view of the EKC we are mostly interested in deterministic trends that do not decay to zero over time and can take $\theta_L = 0$. We can then use simulated inference with $N = T$.

4.3.3 KPSS-type test for the null of cointegration

The correct specification of the nonlinear cointegrating relation will result in a stationary error process $\{u_t\}_{t \in \mathbb{Z}}$. Stationarity tests can thus be used to detect spurious relationships and/or the omission of relevant terms from the

cointegrating regression. We consider a KPSS-type test statistic for the null hypothesis of stationarity. The test statistic reads

$$K_T^+ = \widehat{\Omega}_{u,v}^{-1} \frac{1}{T} \sum_{t=1}^T \left(\frac{1}{\sqrt{T}} \sum_{i=1}^T \hat{u}_i^+ \right)^2, \quad (4.7)$$

where $\widehat{\Omega}_{u,v}$ is a consistent estimator of $\Omega_{u,v} = \Omega_{uu} - \Omega_{uv}\Omega_{vv}^{-1}\Omega_{vu}$, $\hat{u}_t^+ = y_t^+ - \mathbf{d}_t(\widehat{\boldsymbol{\theta}}_T)' \widehat{\boldsymbol{\tau}}_T - \mathbf{s}_t' \widehat{\boldsymbol{\phi}}_T$, and $y_t^+ = y_t - \widehat{\Omega}_{vu} \widehat{\Omega}_{vv}^{-1} \Delta \mathbf{x}_t$. The statistic is (stochastically) bounded under the null hypothesis but diverges under the alternative. Several authors have reported model settings in which the asymptotic null distribution of K_T^+ is known, e.g. Kwiatkowski et al. (1992) and Wagner and Hong (2016).

The estimation of $\boldsymbol{\theta}$ contaminates the limiting distribution of (4.7) with nuisance parameters.⁸ Choi and Saikkonen (2010), Wagner and Hong (2016), Jiang et al. (2019), and Chapter 3, have shown that subsampling can resolve this issue. We will follow their approach and use subsamples of size q_T to compute the test statistics.

Theorem 4.4

Under Assumptions 4.1-4.3 and if $\lim_{T \rightarrow \infty} \left(q_T^{-1} + (\ln T) \left(\frac{q_T}{T} \right)^{\theta_L + \frac{1}{2}} \right) = 0$, then for any $\ell \in \{1, \dots, T - q_T + 1\}$ we have

$$K_{q_T, \ell}^+ = \widehat{\Omega}_{u,v}^{-1} \frac{1}{q_T} \sum_{t=\ell}^{\ell+q_T-1} \left(\frac{1}{\sqrt{q_T}} \sum_{i=\ell}^t \hat{u}_i^+ \right)^2 \Rightarrow \int_0^1 [W(r)]^2 dr, \quad (4.8)$$

where $\widehat{\Omega}_{u,v}$ is a consistent estimator of $\Omega_{u,v} = \Omega_{uu} - \Omega_{uv}\Omega_{vv}^{-1}\Omega_{vu}$ (Theorem 4.2) and $W(\cdot)$ denotes a standard Brownian motion.

Theorem 4.4 does not provide any guidance on the choices for the starting

⁸Proposition 5 in Wagner and Hong (2016) shows that the limiting distribution of K_T^+ is free of nuisance parameters if $\boldsymbol{\theta}_0$ is known and only a single integrated regressor occurs with integer powers greater than one. This result does not carry over to the current setting because of the estimation error in $\widehat{\boldsymbol{\theta}}_T$

value ℓ and the subsample size q_T . First, for a given q_T , Choi and Saikkonen (2010) argue that the use of a single subsample (instead of all T observations) implies a significant loss of power. We follow their example and combine all $M = \lceil T/q_T \rceil$ subresidual series of length q_T using a Bonferroni procedure. That is, we create subresiduals series by selecting adjacent blocks of q_T residuals while alternating between the start and end of the sample. We calculate the KPSS-type test statistic for each subseries, say K_1, \dots, K_M , and reject the null of stationarity at significance α whenever $\max\{K_1, \dots, K_M\}$ exceeds $c_{\alpha/M}$ which is defined by $\mathbb{P}\left(\int_0^1 [W(r)]^2 dr \geq c_{\alpha/M}\right) = \alpha/M$. Finally, we select the block size q_T using Romano and Wolf's (2001) minimum volatility rule. The approach is now completely data-driven.

4.4 Simulations

This section lists various Monte Carlo simulations showing that the asymptotic approximations from Section 4.3 provide useful guidance in finite samples. Further details on the implementation are as follows. We consider $T \in \{100, 200, 500\}$.⁹ The long-run covariance matrices in (4.6) are computed using the Barlett kernel, $k(x) = 1 - |x|$ for $|x| \leq 1$ (and zero otherwise), and the bandwidth selection method described in Andrews (1991). Simulated limiting distributions are based on $J = 999$ replicates and we set $N = T$ (because we consider $\theta_L = 0$ in our settings). We test at 5% significance and report results based on 2.5×10^4 MC replications.

Foreshadowing the empirical application, we use the DGP

$$y_t = \tau_1 + \tau_2 t + \tau_3 t^\theta + \phi_1 x_t + \phi_2 x_t^2 + u_t, \quad (4.9)$$

where $x_t = \sum_{s=1}^t v_s$. The chosen parameter values are $\theta = 2$, $\tau = [\tau_1, \tau_2, \tau_3]' = [7, 0.05, -5 \times 10^{-4}]'$, and $\phi = [\phi_1, \phi_2]'$ and these are representative of typical

⁹Figure 4.1 is an exception. This figure reports power curves and here we decided to use $T \in \{100, 200, 300\}$ instead. This reduces the computational burden and avoids large parameter ranges with empirical power being equal to one.

outcomes in empirical studies. The disturbance vector $[u_t, v_t]'$ is generated from the VAR(1) specification¹⁰

$$\begin{bmatrix} u_t \\ v_t \end{bmatrix} = \mathbf{A} \begin{bmatrix} u_{t-1} \\ v_{t-1} \end{bmatrix} + \begin{bmatrix} \eta_t \\ \epsilon_t \end{bmatrix}, \quad \begin{bmatrix} \eta_t \\ \epsilon_t \end{bmatrix} \stackrel{i.i.d.}{\sim} \mathbf{N} \left(\mathbf{0}, \begin{bmatrix} 1 & \rho \\ \rho & 1 \end{bmatrix} \right). \quad (4.10)$$

In (4.10), we construct the autoregressive matrix \mathbf{A} along the following two steps: (1) generate a (2×2) random matrix \mathbf{U} from $\mathbf{U}[0, 1]$ to construct the orthogonal matrix $\mathbf{H} = \mathbf{U}(\mathbf{U}'\mathbf{U})^{-1/2}$, and (2) compute $\mathbf{A} = \mathbf{H}\mathbf{L}\mathbf{H}'$ using

$$\begin{aligned} \text{(A) } \mathbf{L} &= \text{diag}[0, 0], & \text{(C) } \mathbf{L} &= \text{diag}[0.7, 0.5], \\ \text{(B) } \mathbf{L} &= \text{diag}[0.5, 0.3], & \text{(D) } \mathbf{L} &= \text{diag}[0.9, 0.7]. \end{aligned}$$

Settings (A)–(D) gradually increase the serial correlation in the error processes. The parameter $\rho \in \{0, 0.25, 0.5\}$ governs the amount of endogeneity.

We report three simulation experiments.¹¹ Our first simulation experiments relate to testing the null of *linear* cointegration, i.e. we test $H_0 : \phi_2 = 0$ versus its two-sided alternative $H_a : \phi_2 \neq 0$. The empirical size ($\phi_2 = 0$) is computed using three different estimators: (1) the NLS estimator with simulated critical values as described in Section 4.3.2 (SimNLS); the NLS estimator with simulated critical values and the true value for $\theta_0 = 2$ being provided (SimNLS(θ_0)); and (3) the FMOLS estimator based on $\theta_0 = 2$ (FMOLS(θ_0)). Since the Fully Modified approach is invalid for estimated θ , we do not report such results in the main text. The related simulation outcomes are listed in Panel (A) of Table 4.2. It is clear that simulated critical values improve size control. We point out two other observations. First, we see that the empirical size is rather insensitive to changes in ρ , whereas the introduction of serial correlation makes the test (more) oversized. This behaviour is well-documented in simulation settings where θ is known and

¹⁰We start the VAR recursions from $\begin{bmatrix} u_0 \\ v_0 \end{bmatrix} = \mathbf{0}$ and subsequently use a presample of 50 observations to reduce the influence of these initial values.

¹¹Additional simulation results regarding inference on θ can be found in appendix. More explicitly, we discuss the empirical coverage and average length of confidence intervals for θ there.

restricted to be a natural number, cf. Wagner and Hong (2016) or Chapter 3. Second, we compare the two estimators that are informed about the quadratic deterministic trend: $\text{SimNLS}(\theta_0)$ and $\text{FMOLS}(\theta_0)$. The model is now linear in its parameters and NLS estimation is no longer necessary. That is, we find ourselves in the model specification previously analysed in detail by Wagner and Hong (2016). The comparison of $\text{SimNLS}(\theta_0)$ and $\text{FMOLS}(\theta_0)$ indicates that simulated inference still offers better size control than FMOLS. We thus recommend the applied researcher to use the simulated approach (even if θ is known).

Subsequent simulations are about testing power (Figure 4.1). Overall, the power curves show the expected power gains when sample sizes increases or ϕ_2 moves farther away from zero. These power simulation indicate that the hypothesis $\phi_2 = 0$ is correctly rejected when the nonlinear stochastic trend variable x_t^2 is indeed relevant. Practically speaking, if $\phi_2 \neq 0$ is interpreted as evidence in favour of an EKC, then the test continues to have power also after flexible time trends have been added. All simulations so far make us conclude that we can properly decide on the relevance/irrelevance of x_t^2 (or the GDP variable within an EKC application) within the GPR framework.

Finally, we look at the finite sample properties of the KPSS test (Panel (B) in Table 4.2). Comparing KPSS and $\text{KPSS}(\theta_0)$, we see that knowledge of the true value of θ_0 is beneficial as it always brings the empirical size closer to 5%. This difference aside, our KPSS outcomes are qualitatively similar to the results reported in table 1 of Choi and Saikkonen (2010). That is, the Bonferroni correction leads to conservative tests for up to moderate levels of serial correlation. At high levels of serial correlation (Setting (D)) we approach unit root behaviour and the KPSS tests are oversized.

Table 4.2: (A): The empirical size (in %) of the coefficient test $H_0 : \phi_2 = 0$ versus $H_a : \phi_2 \neq 0$. The Monte Carlo results are based on: simulated inference with θ estimated by NLS (SimNLS), simulated inference with $\theta = 2$ (SimNLS(θ_0)), and a Fully Modified estimator being informed about $\theta = 2$ (FMOLS(θ_0)). (B): The empirical size (in %) of the subsampling Bonferroni KPSS tests. The row labeled ‘KPSS’ is computable in practice. We additionally report simulation outcomes on the same test when being informed about the true value of θ , see the row indicated by ‘KPSS(θ_0)’.

	(A)			(B)			(C)			(D)		
ρ	0.00	0.25	0.50	0.00	0.25	0.50	0.00	0.25	0.50	0.00	0.25	0.50
Panel (A): Empirical size of coefficient test for $H_0 : \phi_2 = 0$ versus $H_a : \phi_2 \neq 0$ (see Figure 4.1 for empirical power)												
$T = 100$												
SimNLS	5.68	5.42	5.38	6.24	5.88	5.81	9.16	9.24	8.70	26.25	26.19	25.61
SimNLS(θ_0)	5.43	5.38	5.31	5.42	5.00	5.08	7.16	7.25	7.01	19.29	19.48	19.28
FMOLS(θ_0)	7.50	7.00	6.58	14.69	14.20	12.90	21.08	20.28	19.12	38.98	38.54	37.88
$T = 200$												
SimNLS	5.44	5.37	5.28	5.12	4.85	5.14	6.66	5.92	5.97	14.44	13.80	13.36
SimNLS(θ_0)	5.50	5.21	5.09	4.77	4.55	4.83	5.54	5.29	5.28	10.00	9.65	9.43
FMOLS(θ_0)	6.08	6.02	5.79	10.90	10.85	10.20	14.58	14.65	13.32	26.02	25.98	25.02
$T = 500$												
SimNLS	5.27	5.27	5.17	4.63	4.72	4.69	4.97	5.02	4.90	6.94	6.72	6.46
SimNLS(θ_0)	5.31	5.34	5.25	4.44	4.67	4.70	4.75	4.74	4.55	5.74	5.71	5.19
FMOLS(θ_0)	5.49	5.38	5.45	8.33	8.40	7.85	10.62	10.08	9.78	14.92	15.22	13.87
Panel (B): Empirical size of the KPSS-type of test for cointegration												
$T = 100$												
KPSS	0.34	0.43	1.34	0.92	1.00	1.39	2.03	2.09	2.16	20.47	20.22	20.22
KPSS(θ_0)	0.69	0.78	1.90	1.32	1.46	1.89	2.30	2.23	2.23	16.48	16.51	16.42
$T = 200$												
KPSS	0.79	0.94	2.31	1.46	1.70	2.42	2.22	2.19	2.26	10.70	10.61	10.17
KPSS(θ_0)	1.19	1.24	2.69	1.82	2.08	2.70	2.43	2.38	2.50	8.92	8.86	8.24
$T = 500$												
KPSS	1.08	1.48	2.90	2.08	1.91	2.40	2.21	2.10	2.43	5.10	4.72	4.20
KPSS(θ_0)	1.36	1.76	3.13	2.27	2.24	2.56	2.32	2.27	2.54	4.35	3.98	3.68

Note: The DGP is $y_t = \tau_1 + \tau_2 t + \tau_3 t^2 + \phi_1 x_t + \phi_2 x_t^2 + u_t$ (numerical values are given on page 170) with $x_t = \sum_{j=1}^d v_{jt}$. The innovations follow $\begin{bmatrix} u_t \\ v_t \end{bmatrix} = A \begin{bmatrix} u_{t-1} \\ v_{t-1} \end{bmatrix} + \begin{bmatrix} u_t \\ v_t \end{bmatrix}$ with $\begin{bmatrix} u_t \\ v_t \end{bmatrix} \stackrel{i.i.d.}{\sim} N\left(0, \begin{bmatrix} 1 & \rho \\ \rho & 1 \end{bmatrix}\right)$. $A = H L H'$ with $H = U(U'U)^{-1/2}$ and U being a (2×2) matrix of uniformly distributed random variables on $[0, 1]$. We consider four specifications for L : (A) $L = \text{diag}[0, 0]$, (B) $L = \text{diag}[0.5, 0.3]$, (C) $L = \text{diag}[0.7, 0.5]$, and (D) $L = \text{diag}[0.9, 0.7]$.

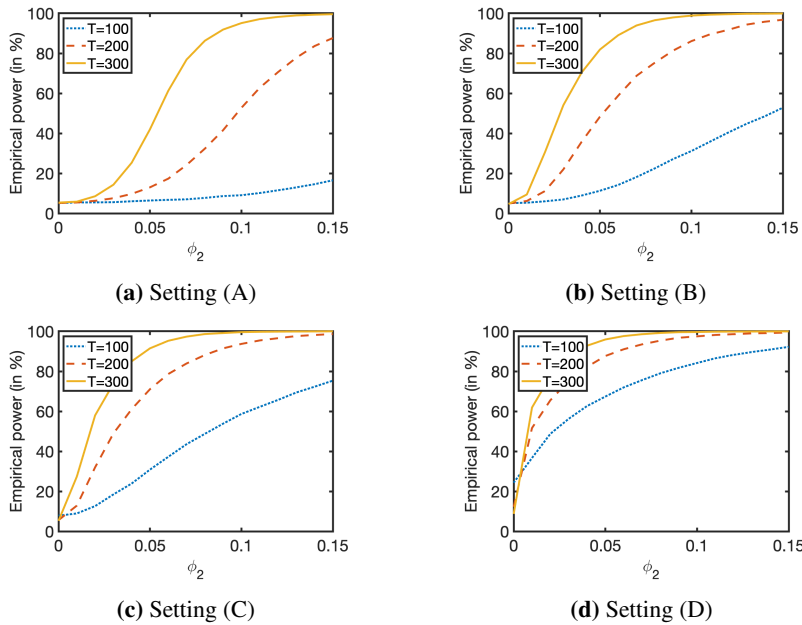


Figure 4.1: The power curve for the test $H_0 : \phi_2 = 0$ versus its two-sided alternative $H_a : \phi_2 \neq 0$. The simulation DGP is $y_t = \tau_1 + \tau_2 t + \tau_3 t^\theta + \phi_1 x_t + \phi_2 x_t^2 + u_t$. Results are shown for settings (A)–(D) with $\rho = 0.50$.

4.5 Empirical Application

We examine the evidence for an EKC for a collection of 18 countries over the period 1870-2014 ($T = 145$). Economic growth is measured by GDP and we use carbon dioxide (CO_2) emissions as a proxy for air pollution. The origin of these data is as follows. We used population and GDP data from the Maddison Project (see <https://www.rug.nl/ggdc/historicaldevelopment/maddison/>). Our carbon dioxide data are fossil-fuel CO_2 emissions as made available by the Carbon Dioxide Information Analysis Center (CDIAC, see <https://cdiac.ess-dive.lbl.gov>). Both GDP and CO_2 emissions are expressed per capita and subsequently log-transformed. In accordance with the notation of this chapter, we will denote them by x_t and y_t , respectively. The same data set (or subsets thereof) has also been studied by Wagner (2015), Chan and Wang (2015), Wang et al. (2018), Wagner et al. (2020), and Chapter 3.¹² This conveniently allows us to compare results. All user choices (kernel specification, bandwidth selection, etc.) are kept the same as during the simulation study (see page 170).

Prior to the analysis of the econometric models we will discuss several features of the time series for Belgium (Figure 4.2).¹³ An inverted U-shaped relationship between GDP and CO_2 (both in log per capita) is clearly visible in Figure 4.2(a) and results like these have triggered the research on the Environmental Kuznetz Curve. However, the heat map time indication also shows that time is almost monotonically increasing along the curve. Time effects - e.g. increasing environmental awareness, advances in sustainable technologies - can be valid alternative explanations for these nonlinearities and their omission can (falsely) exaggerate the influence of GDP. It is for this reason that we developed and analysed the Generalized Cointegrating Polynomial Regression (GCPR).

¹²The stationarity properties of the series have been extensively studied and commented on in these papers. We will not repeat this analysis but refer the interested reader to the appendix where such results can be found.

¹³The data for Austria, Belgium, and Finland are mentioned in both Wagner (2015) and Wagner et al. (2020) to behave in line with the EKC. We discuss Belgium in the main text but the interested reader can find the same figures for Austria and Finland in appendix. The conclusions are the same.

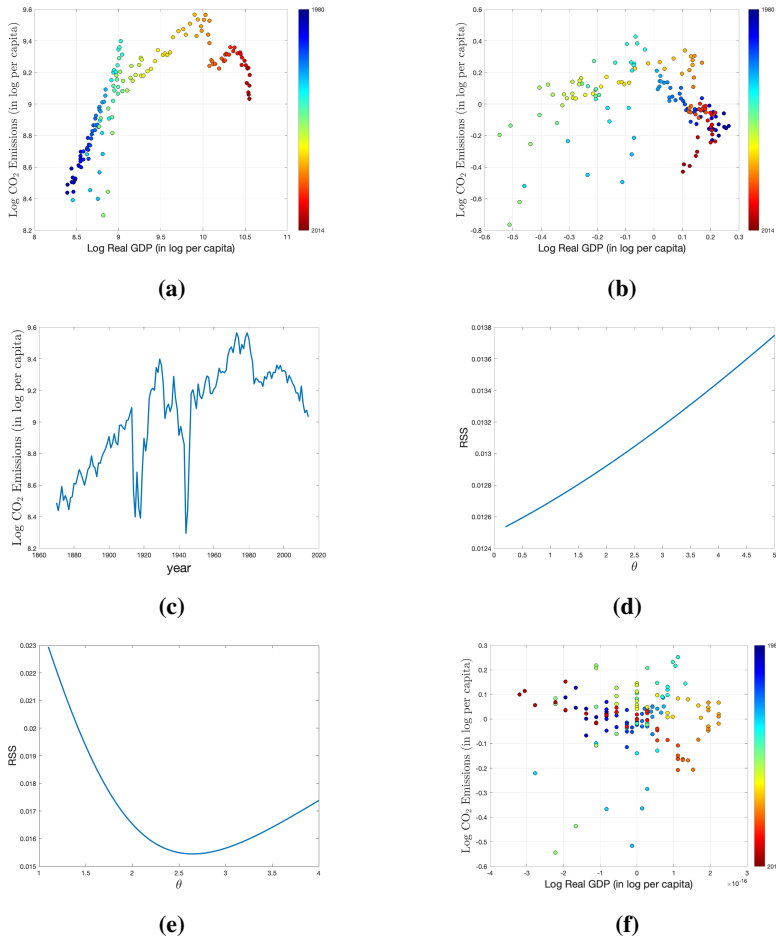


Figure 4.2: Overview graphs for Belgium over 1870-2014. **(a)** $\log(\text{CO}_2)$ versus $\log(\text{GDP})$ (both per capita). **(b)** As subfigure (a) but using detrended variables. **(c)** The log per capita CO₂ emissions time series for Belgium. **(d)** The residual sum of squares (RSS) for the nonlinear model specification $y_t = \tau_1 + \tau_2 t + \phi_1 x_t + \phi_2 x_t^\theta + u_t$ for various values of θ . **(e)** The RSS as a function of θ for the flexible nonlinear trend specification $y_t = \tau_1 + \tau_2 t + \tau_3 t^\theta + \phi x_t + u_t$. **(f)** The relation between x_t and y_t after partialling out the constant, linear trend, and flexible deterministic trend.

Table 4.3: An overview of all models discussed in the empirical application.

Model	Specification
(M1)	$y_t = \tau_1 + \tau_2 t + \phi_1 x_t + \phi_2 x_t^2 + u_t$
(M2)	$y_t = \tau_1 + \tau_2 t + \tau_3 t^2 + \phi_1 x_t + \phi_2 x_t^2 + u_t$
(M3)	$y_t = \tau_1 + \tau_2 t + \tau_3 t^\theta + \phi_1 x_t + \phi_2 x_t^2 + u_t$
(M4)	$y_t = \tau_1 + \tau_2 t + \tau_3 t^\theta + \phi_1 x_t + u_t$

More evidence for the importance of time effects is available in Figure 4.2(b). This figure depicts the same per capita series after detrending.¹⁴ The inverted U-shape is now (visually) less pronounced or even absent.

Finally, we consider two competing possibilities to extend the traditional linear cointegration specification: $y_t = \tau_1 + \tau_2 t + \phi_1 x_t + u_t$. This model does not account for any nonlinear behaviour over time and is therefore ill-suited to fit the data displayed in Figure 4.2(c). Cointegrating polynomial regressions use integer powers of x_t to describe the curvature over time. Following Hu et al. (2019) we can allow for an integrated regressor with a flexible power and estimate $y_t = \tau_1 + \tau_2 t + \phi_1 x_t + \phi_2 x_t^\theta + u_t$. The residual sum of squares (RSS) of the NLS estimator for this specification is shown in Figure 4.2(d). The absence of a minimum at $\theta = 2$ casts doubt on the commonly used quadratic specification in x_t . Moreover, the lack of any minimum might be interpreted as a sign that log per capita GDP is not the source of nonlinearity. Alternatively, we can opt for a flexible deterministic trend as in $y_t = \tau_1 + \tau_2 t + \tau_3 t^\theta + \phi_1 x_t + u_t$. The RSS now exhibits a clear minimum, see Figure 4.2(e). Considerations like these motivate the use of GCPRs. We will argue in the next pages that this last model specification is well-suited to capture the important features of the pollution data.

¹⁴The Perron and Yabu (2009) test allows us to test for the presence of a deterministic trend irrespectively of the series being trend-stationary or having an unit root. The results of this test (see appendix) indicate that log per capita GDP is likely to have a deterministic trend component. It is thus recommended to have a deterministic trend in the model for log per capita CO₂ emissions and the visual inspection of the relationship between GDP and CO₂ emissions (in log per capita) should take place after partialling out this deterministic trend.

We continue the empirical analysis with a comparison of three model specifications, namely (M1)–(M3) as defined in Table 4.3. All three models are of the form:

$$y_t = \tau_1 + \tau_2 t + \tau_3 t^\theta + \phi_1 x_t + \phi_2 x_t^2 + u_t. \quad (4.11)$$

Model (M1) is the specification above with $\tau_3 = 0$. This model specification (possibly with the additional constraint $\tau_2 = 0$) has been explored in various papers, e.g. Piaggio and Padilla (2012), Wagner (2015), and Wang et al. (2018). For this model specification (M1), an inverted-U relationship results when $\phi_1 > 0$ and $\phi_2 < 0$ and empirical evidence hereof is traditionally interpreted as the existence of the EKC. Moreover, if these coefficients have the correct signs, then the turning point - the level of economic growth at which environmental improvement starts - can be computed as $\exp(-\phi_1/2\phi_2)$. Model (M1) is restrictive in the sense that nonlinear time effects (clearly visible in Figure 4.2) can only be explaining using the term $\phi_2 x_t^2$.

The model specifications (M2) and (M3) include deterministic nonlinear time trends. For model (M2), we allow for $\tau_3 \neq 0$ but fix $\theta = 2$. The model in (4.11) without further restrictions is referred to as (M3). In the latter model, the NLS estimator for θ is computed by a grid search over the values $\Theta = [0.05, 0.95] \cup [1.05, 10]$ and simulated inference is used (Section 4.3.2). Table 4.4 depicts how increasingly flexible nonlinear deterministic trends affect the parameter estimates for ϕ_1 and ϕ_2 . Judging only by the signs of $\widehat{\phi}_1$ and $\widehat{\phi}_2$ (thus ignoring potential stationarity in the errors), the EKC exists for 17 out of 18, 9 out of 18, and 8 out of 18 countries for (M1), (M2), and (M3), respectively. Moreover, the significance of squared log per capita GDP (that is ϕ_2) reduces when nonlinear deterministic time trends are included. For model (M3), ϕ_2 is never significantly different from zero at a 10% level and evidence in favour of EKC becomes rather meagre. The results of the KPSS tests for these models can be found in Table 4.4 under “Stationarity tests”. In general, the cointegrating relations seem well-specified except maybe for Belgium, Denmark, and UK.¹⁵

¹⁵Deciding on the correct specification of the cointegrating relations for each of the 18 countries

Table 4.4: Parameter estimates and output of the KPSS-type of test for stationarity as computed for model specifications (M1) – (M4). The column $KPSS$ and M_{opt} provide the numerical values of the KPSS tests and the number of chosen residual subblocks, respectively.

Country	Parameter estimates								Stationarity tests							
	(M1)		(M2)		(M3)		(M4)		(M1)		(M2)		(M3)		(M4)	
	$\hat{\phi}_1$	$\hat{\phi}_2$	$\hat{\phi}_1$	$\hat{\phi}_2$	$\hat{\phi}_1$	$\hat{\phi}_2$	$\hat{\theta}$	$\hat{\phi}_1$	$KPSS$	M_{opt}	$KPSS$	M_{opt}	$KPSS$	M_{opt}	$KPSS$	M_{opt}
Australia	2.75	-0.17	-23.92***	1.40***	-12.19***	0.74	0.88	1.25***	1.49	9	1.79	9	1.80	9	1.35	9
Austria	7.13***	-0.30**	1.25	0.03	3.75***	-0.12	0.88	1.55***	1.03	7	1.07	7	1.65	7	1.63	7
Belgium	11.45***	-0.57***	10.03***	-0.49***	10.29***	-0.50	2.60	1.01***	1.91	9	2.92*	8	2.74*	8	2.28	9
Canada	12.72***	-0.64***	14.80	-0.77	-3.46***	0.25	0.56	1.14***	2.83*	7	2.60*	7	1.26	9	1.28	9
Denmark	14.52***	-0.65***	-2.80	0.25	-5.75***	0.39	2.03	1.68***	3.14*	9	3.30**	9	2.98*	9	1.58	8
Finland	16.86***	-0.76***	16.97***	-0.77***	22.58***	-1.06	1.87	3.95***	2.05	8	2.06	8	0.71	9	0.80	9
France	10.87***	-0.55***	3.14*	-0.12	3.31***	-0.13	2.09	1.00***	1.72	9	0.69	8	0.56	8	2.49	9
Germany	6.24***	-0.31***	-1.82	0.13	-4.42***	0.29	0.59	0.89***	2.63*	7	2.1	8	1.23	9	2.81*	7
Italy	11.76***	-0.55***	7.31**	-0.30	7.72***	-0.29	0.82	2.41***	4.18**	7	3.79**	8	0.79	7	0.78	7
Japan	9.86***	-0.52***	-4.27	0.29	1.16***	0.00	0.05	1.15***	5.17***	8	3.93**	7	1.83	9	1.84	9
Netherlands	8.70***	-0.41***	1.49	-0.01	0.48**	0.05	1.86	1.32***	2.16	7	0.94	7	1.2	7	1.15	7
Norway	3.87	-0.16	-9.14**	0.51**	-1.10**	0.16	0.46	2.05***	2.53*	7	1.06	7	0.74	9	1.44	8
Portugal	0.09	0.04	-5.86***	0.42	-1.11**	0.15	0.05	1.69***	6.95***	8	5.28**	7	0.64	7	1.64	7
Spain	7.72***	-0.37***	1.98	-0.01	4.31***	-0.16	1.55	1.52***	2.78*	7	2.03	8	2.4	8	2.42	8
Sweden	10.91***	-0.44***	-9.08*	0.61**	0.43	0.17	0.46	3.48***	3.59**	7	1.27	7	0.73	7	0.80	7
Switzerland	8.57***	-0.29***	-7.86**	0.54***	-13.86***	0.83	2.98	2.63***	0.80	7	0.96	7	0.77	7	0.75	7
UK	9.32***	-0.47***	5.91***	-0.27***	4.13***	-0.18	3.04	0.80***	2.76*	9	4.25**	9	4.30**	9	3.98**	9
USA	8.67***	-0.44***	0.93	-0.03	-5.62***	0.35	0.92	0.95***	1.64	8	1.85	8	2.25	8	1.97	8

Note: Asterisks denote rejection of the null hypothesis at the *** 1%, ** 5%, and * 10% significance level. Depending on the specific table entry, the null hypothesis refers to either a coefficient being zero or (nonlinear) cointegration.

The insignificance of ϕ_2 in model (M3) suggests us to return to the model specification that was introduced earlier, namely

$$y_t = \tau_1 + \tau_2 t + \tau_3 t^\theta + \phi_1 x_t + u_t. \quad (M4)$$

Model (M4) specifies a linear cointegrating relation around a flexible time trend and does not incorporate nonlinear effects in log per capita GDP.¹⁶ That is, the model specification does not allow for an EKC. As before, we check parameter estimates and test for stationarity of the error terms (the columns labeled “(M4)” in Table 4.4 and Figures 4.2(f), and 4.6). Some remarks concerning this final model specification are:

1. For Belgium, the fitted model reads

$$\hat{y}_t = -0.049 + 0.0063t - 6.131 \times 10^{-6} t^{2.603} + 1.006x_t. \quad (4.12)$$

The flexible power on the linear trend is estimated to be $\hat{\theta} = 2.603$ resulting in nonlinear behaviour over time. Moreover, the negative coefficient in front of $t^{2.603}$ provides a contribution that is sloping down over time. If time effects are ignored, then a 1% increase in GDP will lead to an estimated 1.006% increase in fossil-fuel CO₂ emissions.

2. The outcomes of the KPSS test do not point towards a misspecified cointegrating relation (Table 4.4). The flexible deterministic trend is apparently sufficient to describe the nonlinear behaviour of log per capita CO₂ emissions over time, that is squared log per capita GDP is not needed in the model. Visual proof is found in Figures 4.2(a), 4.2(b) and 4.2(f) where the incorporation of increasingly flexible time effects is seen to remove any apparent nonlinear relationship between log per capita GDP and CO₂ emissions.

is implicitly a joint test. The interpretation of the individual outcomes therefore suffers from the multiple testing problem. A multivariate stationarity test is discussed in Chapter 3.

¹⁶Model specification (M4) has the additional advantage of being invariant to the possible presence of a drift component in log per capita GDP, also see footnote 14.

4.6 Summary and conclusion

In this chapter we have extended the cointegrating polynomial regression model of Wagner and Hong (2016) with power law deterministic trends. The unknown powers are estimated jointly with the parameters in the cointegrating relation. The limiting distribution is nonstandard because it involves a non-diagonal scaling matrix and the usual second order bias effects. We therefore suggest a simulation-based approach to conduct inference. The usual subsampling KPSS-type for stationarity of the innovations of the nonlinear cointegrating relation remains valid. Our results are supported by Monte Carlo simulation. The empirical application on the Environmental Kuznets Curve shows that a flexible trend can fully capture the nonlinearity in the data thereby making higher order powers of log per capita GDP redundant. Our resulting model is linear in log per capita GDP and suggests an alternative explanation in which time effects (e.g. technological progress, environmental awareness) cause the recent slowdown in pollution. Contrary to the opening quote in the introduction, our data does not suggest that air quality will benefit from economic growth.

Appendix 4.A Useful Lemmas

In this section, we first show some preliminary results that will be used in the proofs of main theorems (Section B).

Lemma 4.1

- (i) For $a_L > -1$, we have $\sup_{a \geq a_L} \left| \frac{1}{T} \sum_{t=1}^T \left(\frac{t}{T} \right)^a \right| \leq C$,
- (ii) Under Assumption 4.2, for any $a > -\frac{1}{2}$, and any $k \geq 0$,

$$\mathbb{E} \left(\frac{1}{\sqrt{T}} \sum_{t=1}^T \left(\frac{t}{T} \right)^a (\ln t)^k u_t \right)^2 \leq C (\ln T)^{2k},$$
- (iii) Under Assumption 4.2, for some a_L and a_U such that $-\frac{1}{2} < a_L < a_U < \infty$,
and any $k \geq 0$,
$$\mathbb{E} \left(\sup_{a \in [a_L, a_U]} \left| \frac{1}{\sqrt{T}} \sum_{t=1}^T \left(\frac{t}{T} \right)^a (\ln t)^k u_t \right| \right) \leq C (\ln T)^k,$$
- (iv) If a_L and a_U satisfy $-1 < a_L < a_U < \infty$, and if $k = 0, 1, 2, \dots$, then

$$\sup_{a \in [a_L, a_U]} \left| \frac{1}{T} \sum_{t=1}^T \left(\frac{t}{T} \right)^a \left(\ln \frac{t}{T} \right)^k - \int_0^1 r^a (\ln r)^k dr \right| \leq C \frac{(\ln T)^{k+1}}{T^{1+\min(a_L, 0)}}.$$

Proof (i) This is shown in lemma 4 of Robinson (2012). (ii) Note that

$$\begin{aligned} \mathbb{E} \left(\frac{1}{\sqrt{T}} \sum_{t=1}^T \left(\frac{t}{T} \right)^a (\ln t)^k u_t \right)^2 &= \frac{1}{T} \sum_{t=1}^T \sum_{s=1}^T \left(\frac{t}{T} \right)^a \left(\frac{s}{T} \right)^a (\ln t)^k (\ln s)^k \mathbb{E}(u_t u_s) \\ &\leq \frac{(\ln T)^{2k}}{T} \sum_{t=1}^T \sum_{s=1}^T \left(\frac{t}{T} \right)^a \left(\frac{s}{T} \right)^a |\mathbb{E}(u_t u_s)| \\ &\leq 2 \frac{(\ln T)^{2k}}{T} \sum_{t=1}^T \sum_{s=0}^{t-1} \left(\frac{t}{T} \right)^a \left(\frac{t-s}{T} \right)^a |\gamma_s|, \end{aligned}$$

where $\gamma_s = \mathbb{E}(u_t u_{t-s})$. For the given index ranges, we also have $|t-s| \leq t$ such that

$$\mathbb{E} \left(\frac{1}{\sqrt{T}} \sum_{t=1}^T \left(\frac{t}{T} \right)^a (\ln t)^k u_t \right)^2 \leq 2 (\ln T)^{2k} \frac{1}{T} \sum_{t=1}^T \left(\frac{t}{T} \right)^{2a} \sum_{s=0}^{\infty} |\gamma_s|. \quad (4.13)$$

The first summation in the RHS of (4.13) is bounded in view of Lemma 4.1(i) and $\sum_{s=0}^{\infty} |\gamma_s| < \infty$ due to Assumption 4.2(a) (cf. Appendix 3.A. in Hamilton (1994)). (iii) Using $\frac{t}{T} = \sum_{s=0}^{t-1} \left[\left(\frac{s+1}{T} \right)^a - \left(\frac{s}{T} \right)^a \right]$ and a change in the order of summation,

$$\begin{aligned}
\sum_{t=1}^T \left(\frac{t}{T} \right)^a (\ln t)^k u_t &= \sum_{t=1}^T \sum_{s=0}^{t-1} \left[\left(\frac{s+1}{T} \right)^a - \left(\frac{s}{T} \right)^a \right] (\ln t)^k u_t \\
&= \sum_{s=0}^{T-1} \left[\left(\frac{s+1}{T} \right)^a - \left(\frac{s}{T} \right)^a \right] \sum_{t=s+1}^T (\ln t)^k u_t \\
&= \left(\frac{1}{T} \right)^a \sum_{t=1}^T (\ln t)^k u_t + \sum_{s=1}^{T-1} \left[\left(\frac{s+1}{T} \right)^a - \left(\frac{s}{T} \right)^a \right] \left(\sum_{t=1}^T (\ln t)^k u_t - \sum_{t=1}^s (\ln t)^k u_t \right) \\
&= \left(\frac{1}{T} \right)^a \sum_{t=1}^T (\ln t)^k u_t + \sum_{t=1}^T (\ln t)^k u_t \\
&\quad - \left(\frac{1}{T} \right)^a \sum_{t=1}^T (\ln t)^k u_t - \sum_{s=1}^{T-1} \left[\left(\frac{s+1}{T} \right)^a - \left(\frac{s}{T} \right)^a \right] \sum_{t=1}^s (\ln t)^k u_t,
\end{aligned}$$

and hence

$$\begin{aligned}
\mathbb{E} \left(\sup_{a \in [a_L, a_U]} \left| \frac{1}{\sqrt{T}} \sum_{t=1}^T \left(\frac{t}{T} \right)^a (\ln t)^k u_t \right| \right) &\leq \mathbb{E} \left| \frac{1}{\sqrt{T}} \sum_{t=1}^T (\ln t)^k u_t \right| \\
&\quad + \mathbb{E} \left(\sup_{a \in [a_L, a_U]} \left| \frac{1}{\sqrt{T}} \sum_{s=1}^{T-1} \left[\left(\frac{s+1}{T} \right)^a - \left(\frac{s}{T} \right)^a \right] \sum_{t=1}^s (\ln t)^k u_t \right| \right). \quad (4.14)
\end{aligned}$$

For the first term in the RHS of (4.14), we have $\mathbb{E} \left| \frac{1}{\sqrt{T}} \sum_{t=1}^T (\ln t)^k u_t \right| \leq \left(\mathbb{E} \left(\frac{1}{\sqrt{T}} \sum_{t=1}^T (\ln t)^k u_t \right)^2 \right)^{1/2} \leq C(\ln T)^k$ by Lemma 4.1(ii) with $a = 0$. For the second term, note that

$$\begin{aligned}
\left| \frac{1}{\sqrt{T}} \sum_{s=1}^{T-1} \left[\left(\frac{s+1}{T} \right)^a - \left(\frac{s}{T} \right)^a \right] \sum_{t=1}^s (\ln t)^k u_t \right| \\
\leq \frac{1}{\sqrt{T}} \sum_{s=1}^{T-1} \left(\frac{s}{T} \right)^a \left| \left(1 + \frac{1}{s} \right)^a - 1 \right| \left| \sum_{t=1}^s (\ln t)^k u_t \right|. \quad (4.15)
\end{aligned}$$

To deal with the supremum of $\left| \left(1 + \frac{1}{s} \right)^a - 1 \right|$ over $[a_L, a_U]$, we define $g_a(x) = (1+x)^a - 1$ for $0 \leq x \leq 1$. If $-\frac{1}{2} < a \leq 1$, then $|g_a(x)| \leq |a|x$ by Bernoulli's inequality. If $a \geq 1$,

then convexity of $g_a(x)$ implies $g_a(x) \leq (1-x)g_a(0) + xg_a(1) \leq (2^a - 1)x$. We conclude that $|g_a(x)| \leq Cx$ for all $a_L \leq a \leq a_U$ and $x \in [0, 1]$. Combining this result with (4.15),

$$\begin{aligned}
 & \mathbb{E} \left(\sup_{a \in [a_L, a_U]} \left| \frac{1}{\sqrt{T}} \sum_{s=1}^{T-1} \left[\left(\frac{s+1}{T} \right)^a - \left(\frac{s}{T} \right)^a \right] \sum_{t=1}^s (\ln t)^k u_t \right| \right) \\
 & \leq \mathbb{E} \left(\frac{1}{\sqrt{T}} \sum_{s=1}^{T-1} \left(\frac{s}{T} \right)^{a_L} \sup_{a \in [a_L, a_U]} \left| \left(1 + \frac{1}{s} \right)^a - 1 \right| \left| \sum_{t=1}^s (\ln t)^k u_t \right| \right) \\
 & \leq C \frac{1}{\sqrt{T}} \sum_{s=1}^{T-1} \left(\frac{s}{T} \right)^{a_L} \frac{1}{s} \mathbb{E} \left| \sum_{t=1}^s (\ln t)^k u_t \right| \\
 & \leq CT^{-(a_L+1/2)} \sum_{s=1}^{T-1} s^{a_L-1/2} (\ln s)^k \\
 & \leq C(\ln T)^k \left[\frac{1}{T} \sum_{s=1}^T \left(\frac{s}{T} \right)^{a_L-1/2} \right] \leq C(\ln T)^k,
 \end{aligned}$$

where we used $\mathbb{E} \left| \sum_{t=1}^s (\ln t)^k u_t \right| \leq \left(\mathbb{E} \left(\sum_{t=1}^s (\ln t)^k u_t \right)^2 \right)^{1/2} \leq Cs^{1/2} (\ln s)^k$ (the steps in the proof of (ii) require a small modification to establish this) to go to the last line, and (i) to obtain the final inequality. The proof is complete since we have bounded both terms in the RHS of (4.14). (iv) If we divide the integral into integration intervals of width $\frac{1}{T}$, then we find

$$\begin{aligned}
 & \sup_{a \in [a_L, a_U]} \left| \frac{1}{T} \sum_{t=1}^T \left(\frac{t}{T} \right)^a \left(\ln \frac{t}{T} \right)^k - \int_0^1 r^a (\ln r)^k dr \right| \\
 & = \sup_{a \in [a_L, a_U]} \left| \sum_{t=1}^T \int_{(t-1)/T}^{t/T} \left(\frac{t}{T} \right)^a \left(\ln \frac{t}{T} \right)^k dr - \sum_{t=1}^T \int_{(t-1)/T}^{t/T} r^a (\ln r)^k dr \right| \\
 & = \sup_{a \in [a_L, a_U]} \left| \frac{1}{T} \left(\frac{1}{T} \right)^a \left(\ln \frac{1}{T} \right)^k - \int_0^{1/T} r^a (\ln r)^k dr \right. \\
 & \quad \left. + \sum_{t=2}^T \int_{(t-1)/T}^{t/T} \left[\left(\frac{t}{T} \right)^a \left(\ln \frac{t}{T} \right)^k - r^a (\ln r)^k \right] dr \right| \\
 & \leq \sup_{a \in [a_L, a_U]} \left| \left(\frac{1}{T} \right)^{a+1} \left(\ln \frac{1}{T} \right)^k \right| + \sup_{a \in [a_L, a_U]} \left| \int_0^{1/T} r^a (\ln r)^k dr \right| \\
 & \quad + \sup_{a \in [a_L, a_U]} \sum_{t=2}^T \int_{(t-1)/T}^{t/T} \left| \left(\frac{t}{T} \right)^a \left(\ln \frac{t}{T} \right)^k - r^a (\ln r)^k \right| dr =: Ia + Ib + Ic, \quad (4.16)
 \end{aligned}$$

using the triangle inequality. Clearly, Ia is bounded by $T^{-(a_L+1)}(\ln T)^k$. For Ib we can use the standard integral (cf. Adams and Essex (2016)), namely $\int_0^{1/T} r^a (\ln r)^k dr = \frac{(-1)^k}{a+1} \left(\frac{1}{T}\right)^{a+1} (\ln T)^k - \frac{k}{a+1} \int_0^{1/T} r^a (\ln r)^{k-1} dr$ for $k \neq -1$, to obtain

$$\begin{aligned} \int_0^{1/T} r^a (\ln r)^k dr &= (-1)^k \left(\frac{1}{T}\right)^{a+1} \sum_{j=0}^{k-1} \frac{k!}{(k-j)!} \frac{1}{(a+1)^{1+j}} (\ln T)^{k-j} \\ &+ (-1)^k \frac{k!}{(a+1)^k} \int_0^{1/T} r^a dr = (-1)^k \left(\frac{1}{T}\right)^{a+1} \sum_{j=0}^k \frac{k!}{(k-j)!} \frac{1}{(a+1)^{1+j}} (\ln T)^{k-j}. \end{aligned}$$

We therefore conclude that

$$\begin{aligned} Ib &\leq \sum_{j=1}^k \frac{k!}{(k-j)!} \sup_{a \in [a_L, a_U]} \frac{1}{(a+1)^{1+j}} \left(\frac{1}{T}\right)^{a+1} (\ln T)^{k-j} \\ &\leq \sum_{j=1}^k \frac{k!}{(k-j)!} \frac{1}{(a_L+1)^{1+j}} \left(\frac{1}{T}\right)^{a_L+1} (\ln T)^{k-j} \leq CT^{-(a_L+1)} (\ln T)^k. \end{aligned}$$

It remains to bound the term Ic . Changing the integration variable to $r = \frac{t}{T} - s$ yields

$$Ic = \sup_{a \in [a_L, a_U]} \sum_{t=2}^T \int_0^{1/T} \left| \left(\frac{t}{T}\right)^a \left(\ln \frac{t}{T}\right)^k - \left(\frac{t}{T} - s\right)^a \left[\ln \left(\frac{t}{T} - s\right)\right]^k \right| ds. \quad (4.17)$$

We subsequently derive an upper bound for the integrand using an approach which mimics the derivations in (D.14) and (D.15) in Robinson (2012). For any $\frac{2}{T} \leq \ell \leq 1$ (such that $0 < s/\ell \leq \frac{1}{2}$), we have

$$\begin{aligned} &|\ell^a (\ln \ell)^k - (\ell - s)^a (\ln(\ell - s))^k| \\ &= |[\ell^a - (\ell - s)^a] (\ln \ell)^k + (\ell - s)^a [(\ln \ell)^k - (\ln(\ell - s))^k]| \\ &\leq |[\ell^a - (\ell - s)^a] (\ln \ell)^k| + |(\ell - s)^a [(\ln \ell)^k - (\ln(\ell - s))^k]| \\ &= \ell^a |\ln \ell|^k |1 - (1 - s/\ell)^a| + \ell^a (1 - s/\ell)^a |(\ln \ell)^k - (\ln(\ell - s))^k| =: IIa + IIb, \quad (4.18) \end{aligned}$$

by the triangle inequality and the fact that $|(\ell - s)^a| = (\ell - s)^a$. For IIa similar arguments as those found below (4.15) give $|1 - (1 - x)^a| \leq Cx$, and hence

$$\begin{aligned} IIa &\leq C\ell^{a_L} |\ln \ell|^k \frac{s}{\ell} \leq C\ell^{a_L-1} |\ln \ell|^k s \\ &\leq C\ell^{a_L-1} |\ln \ell|^k T^{-1} \leq C\ell^{a_L-1} |\ln \ell|^k T^{-1} \leq C\ell^{a_L-1} (\ln T)^k T^{-1}, \quad (4.19) \end{aligned}$$

since $|\ln \ell| \leq |\ln T|$ for all $\frac{2}{T} \leq \ell \leq 1$. For *Iib* we first note that $\frac{1}{2} \leq 1 - s/\ell < 1$ and therefore $(1 - s/\ell)^a < (1 - s/\ell)^{-1} \leq 2$. Moreover, we use the factorization $p^n - q^n = (p - q) \sum_{j=0}^{n-1} p^{n-1-j} q^j$ to obtain¹⁷

$$\begin{aligned} |(\ln \ell)^k - (\ln(\ell - s))^k| &= |\ln(1 - s/\ell)| \left| \sum_{j=0}^{k-1} (\ln \ell)^{k-1-j} (\ln(\ell - s))^j \right| \\ &\leq |\ln(1 - s/\ell)| \sum_{j=0}^{k-1} |\ln \ell|^{k-1-j} |\ln(\ell - s)|^j \leq k |\ln(1 - s/\ell)| (\ln T)^{k-1} \leq 2k \frac{s}{\ell} (\ln T)^{k-1}, \end{aligned}$$

because $1/T \leq \ell - s < 1$ and thus $|\ln(\ell - s)| \leq \ln T$. Combining all previous results for *Iib* gives

$$Iib \leq C \ell^a \frac{s}{\ell} (\ln T)^{k-1} \leq C \ell^{a_L-1} (\ln T)^{k-1} \frac{1}{T}.$$

Since $\frac{2}{T} \leq \ell \leq 1$, we use the bounds on *Iia* and *Iib* to bound the integrand of (4.17) as follows:

$$Ic \leq C \sup_{a \in [a_L, a_U]} \sum_{t=2}^T \int_0^{1/T} \left(\frac{t}{T} \right)^{a_L-1} \frac{1}{T} (\ln T)^k ds \leq C \frac{(\ln T)^k}{T^2} \sum_{t=1}^T \left(\frac{t}{T} \right)^{a_L-1}. \quad (4.20)$$

The asymptotic order of $\sum_{t=1}^T \left(\frac{t}{T} \right)^{a_L-1}$ relies on the values of a_L . We distinguish three cases: (1) if $a_L < 0$, then $\sum_{t=1}^T \left(\frac{t}{T} \right)^{a_L-1} = T^{1-a_L} \sum_{t=1}^T \frac{1}{t^{1-a_L}} = T^{1-a_L} O(1)$, (2) if $a_L = 0$, then $\sum_{t=1}^T \left(\frac{t}{T} \right)^{a_L-1} = T \sum_{t=1}^T t^{-1} = T O(\ln T)$, and (3) if $a_L > 0$, $\sum_{t=1}^T \left(\frac{t}{T} \right)^{a_L-1} = O(T)$ by Lemma 4.1(i). By (4.20),

$$Ic = O \left(\frac{(\ln T)^k}{T^{a_L+1}} \mathbf{1}_{\{a_L < 0\}} + \frac{(\ln T)^{k+1}}{T} \mathbf{1}_{\{a_L = 0\}} + \frac{(\ln T)^k}{T} \mathbf{1}_{\{a_L > 0\}} \right). \quad (4.21)$$

We find *Ia*, *Ib*, and *Ic* converge to zero as $T \rightarrow \infty$. The proof follows from (4.16). ■

¹⁷For any $x > -1$, we have the inequality $\frac{x}{1+x} \leq \ln(1+x) \leq x$. This implies that $|\ln(1 - s/\ell)| = -\ln(1 - s/\ell) \leq \frac{s/\ell}{1-s/\ell} \leq 2 \frac{s}{\ell}$.

Lemma 4.2

Let Assumption 4.2 hold. For any a such that $-\frac{1}{2} < a_L \leq a \leq a_U < \infty$, any $i \in \{1, 2, \dots, m\}$, any $j \in \{1, 2, \dots, p_i\}$, and $k \in \{0, 1, 2, \dots\}$, we have:

$$(i) \quad \frac{1}{\sqrt{T}} \sum_{t=1}^T \left(\frac{x_{it}}{\sqrt{T}} \right)^j u_t \Rightarrow \int_0^1 B_{v_i}^j(r) dB_u(r) + j \Delta_{v_i u} \int_0^1 B_{v_i}^{j-1}(r) dr,$$

$$(ii) \quad \frac{1}{\sqrt{T}} \sum_{t=1}^T \left(\frac{t}{T} \right)^a \left(\ln \frac{t}{T} \right)^k u_t \Rightarrow \int_0^1 r^a (\ln r)^k dB_u(r),$$

$$(iii) \quad \frac{1}{T} \sum_{t=1}^T \left(\frac{t}{T} \right)^a \left(\ln \frac{t}{T} \right)^k \left(\frac{x_{it}}{\sqrt{T}} \right)^j \Rightarrow \int_0^1 r^a (\ln r)^k B_{v_i}^j(r) dr.$$

Proof For $r \in (0, 1]$, we define $f(r) = r^a (\ln r)^k$. Two partial sum processes are defined as $S_T(r) = \frac{1}{\sqrt{T}} \sum_{s=1}^{\lfloor rT \rfloor} u_s$, and $X_{i,T}(r) = \frac{1}{\sqrt{T}} x_{i,\lfloor rT \rfloor} = \frac{1}{\sqrt{T}} \sum_{i=1}^{\lfloor rT \rfloor} v_{it}$. Finally, set $f_T(r) = \left(\frac{\lfloor rT \rfloor}{T} \right)^a \left(\ln \frac{\lfloor rT \rfloor}{T} \right)^k$ for $r \in \left[\frac{1}{T}, 1 \right]$. (i) This result follows from lemma 1 of Hong and Phillips (2010). (ii) We have

$$\begin{aligned} \frac{1}{\sqrt{T}} \left(\frac{t}{T} \right)^a \left(\ln \frac{t}{T} \right)^k u_t &= f_T \left(\frac{t}{T} \right) \frac{u_t}{\sqrt{T}} = f_T \left(\frac{t}{T} \right) \left[S_T \left(\frac{t}{T} \right) - S_T \left(\frac{t-1}{T} \right) \right] \\ &= \left[f_T \left(\frac{t}{T} \right) S_T \left(\frac{t}{T} \right) - f_T \left(\frac{t-1}{T} \right) S_T \left(\frac{t-1}{T} \right) \right] - \left[f_T \left(\frac{t}{T} \right) - f_T \left(\frac{t-1}{T} \right) \right] S_T \left(\frac{t-1}{T} \right) \end{aligned} \quad (4.22)$$

and hence

$$\begin{aligned} \frac{1}{\sqrt{T}} \sum_{t=1}^T \left(\frac{t}{T} \right)^a \left(\ln \frac{t}{T} \right)^k u_t &= \left(\frac{1}{T} \right)^a \left(\ln \frac{1}{T} \right)^k \frac{u_1}{\sqrt{T}} + \frac{1}{\sqrt{T}} \sum_{t=2}^T \left(\frac{t}{T} \right)^a \left(\ln \frac{t}{T} \right)^k u_t \\ &\stackrel{(4.22)}{=} f_T \left(\frac{1}{T} \right) S_T \left(\frac{1}{T} \right) + \left[f_T(1) S_T(1) - f_T \left(\frac{1}{T} \right) S_T \left(\frac{1}{T} \right) \right] - \sum_{t=2}^T \left[f_T \left(\frac{t}{T} \right) - f_T \left(\frac{t-1}{T} \right) \right] S_T \left(\frac{t-1}{T} \right) \\ &\stackrel{f_T(1)=0}{=} - \sum_{t=2}^T \int_{(t-1)/T}^{t/T} S_T(r) df_T(r) \end{aligned} \quad (4.23)$$

where we used the fact that $S_T(\cdot)$ is piecewise constant. In view of Assumption 4.2, we can suitably extend the probability space and have the following uniformly strong approximation of $S_T(\cdot)$ (see for example page 562 of Phillips (2007)):

$$\sup_{1 \leq t \leq T} \left| S_T \left(\frac{t-1}{T} \right) - B_u \left(\frac{t-1}{T} \right) \right| = o_{a.s.} \left(\frac{1}{T^{(1/2)-(1/q)}} \right), \quad (4.24)$$

for $q > 2$. Continuing from (4.23), this uniformly strong approximation gives

$$\begin{aligned}
 \frac{1}{\sqrt{T}} \sum_{t=1}^T \left(\frac{t}{T}\right)^a \left(\ln \frac{t}{T}\right)^k u_t &= - \sum_{t=2}^T \int_{(t-1)/T}^{t/T} B_u(r) df_T(r) + o_{a.s.} \left(\frac{1}{T^{(1/2)-(1/q)}} \right) \\
 &= - \int_{1/T}^1 B_u(r) df_T(r) + o_{a.s.} \left(\frac{1}{T^{(1/2)-(1/p)}} \right) \\
 &= B_u \left(\frac{1}{T} \right) f_T \left(\frac{1}{T} \right) + \int_{1/T}^1 f_T(r) dB_u(r) + o_{a.s.} \left(\frac{1}{T^{(1/2)-(1/p)}} \right) \\
 &= \int_0^1 f(r) dB_u(r) - \int_0^{1/T} f(r) dB_u(r) + B_u \left(\frac{1}{T} \right) f_T \left(\frac{1}{T} \right) \\
 &\quad + \int_{1/T}^1 [f_T(r) - f(r)] dB_u(r) + o_{a.s.} \left(\frac{1}{T^{(1/2)-(1/p)}} \right),
 \end{aligned}$$

where the third line is obtained using integration by parts of the mean square Riemann-Stieltjes integral, c.f. theorem 2.7 in Tanaka (2017). It remains to show that $\int_0^{1/T} f(r) dB_u(r)$, $B_u \left(\frac{1}{T} \right) f_T \left(\frac{1}{T} \right)$, and $\int_{1/T}^1 [f_T(r) - f(r)] dB_u(r)$ are asymptotically negligible. These quantities are zero mean so it suffices to show that their variances vanish as $T \rightarrow \infty$. By the isometry property and steps similar to those above (4.17),

$$\mathbb{V}\text{ar} \left(\int_0^{1/T} f(r) dB_u(r) \right) = \Omega_{uu} \int_0^{1/T} [f(r)]^2 dr \leq CT^{-(2a_L+1)} (\ln T)^{2k} \rightarrow 0, \quad (4.25)$$

as $T \rightarrow \infty$. Also, $\mathbb{V}\text{ar} \left(B_u \left(\frac{1}{T} \right) f_T \left(\frac{1}{T} \right) \right) = \frac{1}{T} \Omega_{uu} \left[f_T \left(\frac{1}{T} \right) \right]^2 = \Omega_{uu} \left(\frac{1}{T} \right)^{2a_L+1} \left(\ln \frac{1}{T} \right)^{2k} \rightarrow 0$.

To control the variance of $\int_{1/T}^1 [f_T(r) - f(r)] dB_u(r)$, we look at

$$\begin{aligned}
 \int_{1/T}^1 |f(r) - f_T(r)|^2 dr &= \sum_{t=2}^T \int_{(t-1)/T}^{t/T} \left| f(r) - \left(\frac{t-1}{T} \right)^a \left(\ln \frac{t-1}{T} \right)^k \right|^2 dr \\
 &= \sum_{t=1}^{T-1} \int_{t/T}^{(t+1)/T} \left| r^a (\ln r)^k - \left(\frac{t}{T} \right)^a \left(\ln \frac{t}{T} \right)^k \right|^2 dr \\
 &= \sum_{t=1}^{T-1} \int_0^{1/T} \left| \left(\frac{t}{T} + s \right)^a \left[\ln \left(\frac{t}{T} + s \right) \right]^k - \left(\frac{t}{T} \right)^a \left(\ln \frac{t}{T} \right)^k \right|^2 ds. \quad (4.26)
 \end{aligned}$$

Now let $\ell \in \left\{ \frac{1}{T}, \frac{2}{T}, \dots, 1 \right\}$ and recall that $0 \leq s \leq \frac{1}{T}$ (hence also $0 \leq \frac{s}{\ell} \leq 1$). Using the

triangle inequality, the expression in absolute values can be bounded as

$$\begin{aligned}
& |(\ell + s)^a (\ln(\ell + s))^k - \ell^a (\ln \ell)^k| \\
&= \left| [(\ell + s)^a - \ell^a] (\ln(\ell + s))^k + \ell^a [(\ln(\ell + s))^k - (\ln \ell)^k] \right| \\
&\leq \left| [(\ell + s)^a - \ell^a] (\ln(\ell + s))^k \right| + \left| \ell^a [(\ln(\ell + s))^k - (\ln \ell)^k] \right| \\
&= \ell^a \left| (1 + s/\ell)^a - 1 \right| |\ln(\ell + s)|^k + \ell^a \left| (\ln(\ell + s))^k - (\ln \ell)^k \right| \\
&=: IIc + IId. \tag{4.27}
\end{aligned}$$

By the inequality $|g_a(x)| \leq Cx$ below (4.15) and the fact that $|\ln(\ell + s)| \leq |\ln \ell| + |\ln(1 + s/\ell)| \leq \ln T + s/\ell$, we obtain $IIc \leq C\ell^{a_L} \frac{s}{\ell} \left| \ln T + \frac{s}{\ell} \right|^k \leq C\ell^{a_L-1} (\ln T)^k \frac{1}{T}$. Moreover, the factorisation $p^n - q^n = (p - q) \sum_{j=0}^{n-1} p^{n-1-j} q^j$ yields

$$\begin{aligned}
IId &= \ell^a |\ln(1 + s/\ell)| \left| \sum_{j=0}^{k-1} (\ln(\ell + s))^{k-1-j} (\ln \ell)^j \right| \\
&\leq k\ell^{a_L} \frac{s}{\ell} |(\ln T) + 1|^{k-1} \leq C\ell^{a_L-1} (\ln T)^{k-1} T^{-1}. \tag{4.28}
\end{aligned}$$

By combination of the bounds on IIc and IId , we conclude that $|(\ell + s)^a (\ln(\ell + s))^k - \ell^a (\ln \ell)^k| \leq C\ell^{a_L-1} (\ln T)^k \frac{1}{T}$ and arrive at the following upper bound on the RHS of (4.26):

$$\begin{aligned}
\int_{1/T}^1 |f(r) - f_T(r)|^2 dr &\leq C(\ln T)^{2k} \frac{1}{T^3} \sum_{t=1}^T \left(\frac{t}{T} \right)^{2(a_L-1)} \\
&= O\left(\frac{(\ln T)^{2k}}{T^{2(a_L+\frac{1}{2})}} \mathbf{1}_{\{a_L < \frac{1}{2}\}} + \frac{(\ln T)^{2k+1}}{T^2} \mathbf{1}_{\{a_L = \frac{1}{2}\}} + \frac{(\ln T)^{2k}}{T^2} \mathbf{1}_{\{a_L > \frac{1}{2}\}} \right). \tag{4.29}
\end{aligned}$$

The RHS of (4.29) will go to zero as $T \rightarrow \infty$, thereby establishing that $\int_{1/T}^1 [f_T(r) - f(r)] dB_u(r)$ is also asymptotically negligible. The proof of part (ii) is now complete.

(iii) We have

$$\begin{aligned}
\frac{1}{T} \sum_{t=1}^T \left(\frac{t}{T} \right)^a \left(\ln \frac{t}{T} \right)^k \left(\frac{x_{it}}{\sqrt{T}} \right)^j &= \sum_{t=2}^T \int_{(t-1)/T}^{t/T} f_T(r) X_{i,T}^j(r) dr \\
&= \int_0^1 f(r) X_{i,T}^j(r) dr + \int_{1/T}^1 [f_T(r) - f(r)] X_{i,T}^j(r) dr =: IIIa + IIIb. \tag{4.30}
\end{aligned}$$

Given the CMT and $X_{i,T} \Rightarrow B_{v_i}$, term $IIIa$ will converge weakly to $\int_0^1 f(r) B_{v_i}^j(r) dr$

if we can show that $x \mapsto \int_0^1 f(r)x^j(r)dr$ is a continuous functional. Let $x, y \in D[0, 1]$. Hölder's inequality implies

$$\begin{aligned} \left| \int_0^1 f(r)x^j(r)dr - \int_0^1 f(r)y^j(r)dr \right| &= \left| \int_0^1 f(r)(x^j(r) - y^j(r))dr \right| \\ &\leq \int_0^1 |f(r)|dr \sup_{r \in [0,1]} |x^j(r) - y^j(r)| \leq C \sup_{r \in [0,1]} |x(r) - y(r)| \rightarrow 0, \end{aligned} \quad (4.31)$$

because $\int_0^1 |f(r)|dr = \frac{k!}{(1+a)^{k+1}}$ is bounded. Continuity of the functional now follows from (4.31). If we apply the Cauchy-Schwartz inequality to *IIIb*, then we find

$$IIIb \leq \left[\int_{1/T}^1 |f(r) - f_T(r)|^2 dr \right]^{1/2} \left[\int_{1/T}^1 X_{i,T}^{2j}(r)dr \right]^{1/2}.$$

Since $\int_{1/T}^1 |f(r) - f_T(r)|^2 dr = o(1)$ by (4.29) and $\int_{1/T}^1 X_{i,T}^{2j}(r)dr = \int_0^1 X_{i,T}^{2j}(r)dr \Rightarrow \int_0^1 B_{v_i}^{2j}(r)dr$. We conclude that $IIIb = o_p(1)$. Now combine the limiting results for *IIIa* and *IIIb* to complete the argument. ■

Lemma 4.3

Let $f(\mathbf{w}_t, \gamma) = \mathbf{d}_t(\boldsymbol{\theta})' \boldsymbol{\tau} + \mathbf{s}_t' \boldsymbol{\phi}$, where $\mathbf{w}_t = [t, \mathbf{x}_t]'$ and $\mathbf{x}_t = [x_{1t}, x_{2t}, \dots, x_{mt}]'$ stacks all the stochastic trends. Let $\dot{f}(\mathbf{w}_t, \gamma)$ and $\ddot{f}(\mathbf{w}_t, \gamma)$ denote the first and second derivatives of $f(\mathbf{w}_t, \gamma)$ with respect to γ .

$$(i) \quad \dot{f}(\mathbf{w}_t, \gamma) = [(\boldsymbol{\tau} \odot \mathbf{d}_t(\boldsymbol{\theta}))' \ln t, \mathbf{d}_t(\boldsymbol{\theta})', \mathbf{s}_t']'.$$

$$(ii) \quad L_{\tau_0, T} \dot{f}(\mathbf{w}_t, \gamma) = \begin{bmatrix} [(\boldsymbol{\tau} - \tau_0) \odot \mathbf{d}_t(\boldsymbol{\theta}))' \ln t + [\tau_0 \odot \mathbf{d}_t(\boldsymbol{\theta}))' \ln \frac{t}{T}] \\ \mathbf{d}_t(\boldsymbol{\theta})' \\ \mathbf{s}_t' \end{bmatrix} \text{ and thus } L_{\tau_0, T} \dot{f}(\mathbf{w}_t, \gamma_0) = \begin{bmatrix} [\tau_0 \odot \mathbf{d}_t(\boldsymbol{\theta}_0)]' \ln \frac{t}{T} \\ \mathbf{d}_t(\boldsymbol{\theta}_0)' \\ \mathbf{s}_t' \end{bmatrix}.$$

(iii) We have

$$\dot{f}(\mathbf{w}_t, \gamma) - \dot{f}(\mathbf{w}_t, \gamma_0) = \begin{bmatrix} ((\boldsymbol{\tau} - \tau_0) \odot (\mathbf{d}_t(\boldsymbol{\theta}) - \mathbf{d}_t(\boldsymbol{\theta}_0)) + \tau_0 \odot (\mathbf{d}_t(\boldsymbol{\theta}) - \mathbf{d}_t(\boldsymbol{\theta}_0)) + (\boldsymbol{\tau} - \tau_0) \odot \mathbf{d}_t(\boldsymbol{\theta}_0))' \ln t \\ \mathbf{d}_t(\boldsymbol{\theta}) - \mathbf{d}_t(\boldsymbol{\theta}_0) \\ \mathbf{0}_{p \times 1} \end{bmatrix}.$$

which implies

$$\begin{aligned} L_{\tau_0, T} (\dot{f}(\mathbf{w}_t, \gamma) - \dot{f}(\mathbf{w}_t, \gamma_0)) \\ = \begin{bmatrix} ((\boldsymbol{\tau} - \tau_0) \odot (\mathbf{d}_t(\boldsymbol{\theta}) - \mathbf{d}_t(\boldsymbol{\theta}_0)) + (\boldsymbol{\tau} - \tau_0) \odot \mathbf{d}_t(\boldsymbol{\theta}_0))' \ln t + \tau_0 \odot (\mathbf{d}_t(\boldsymbol{\theta}) - \mathbf{d}_t(\boldsymbol{\theta}_0))' \ln \frac{t}{T} \\ \mathbf{d}_t(\boldsymbol{\theta}) - \mathbf{d}_t(\boldsymbol{\theta}_0) \\ \mathbf{0}_{p \times 1} \end{bmatrix}. \quad (4.32) \end{aligned}$$

$$(iv) \quad \ddot{f}(\mathbf{w}_t, \gamma) = \begin{bmatrix} \text{diag}[\boldsymbol{\tau}] \text{diag}[\mathbf{d}_t(\boldsymbol{\theta})] (\ln t)^2 & \text{diag}[\mathbf{d}_t(\boldsymbol{\theta})] \ln t & \mathbf{0}_{d \times p} \\ \text{diag}[\mathbf{d}_t(\boldsymbol{\theta})] \ln t & \mathbf{0}_{d \times d} & \mathbf{0}_{d \times p} \\ \mathbf{0}_{p \times d} & \mathbf{0}_{p \times d} & \mathbf{0}_{p \times p} \end{bmatrix}.$$

$$(v) \quad \mathbf{G}_{\gamma_0, T}^{-1} \ddot{f}(\mathbf{w}_t, \gamma) \mathbf{G}_{\gamma_0, T}^{-1} = \begin{bmatrix} \ddot{F}_{11, t} & \ddot{F}_{12, t} & \mathbf{0}_{d \times p} \\ \ddot{F}_{21, t} & \mathbf{0}_{d \times d} & \mathbf{0}_{d \times p} \\ \mathbf{0}_{p \times d} & \mathbf{0}_{p \times d} & \mathbf{0}_{p \times p} \end{bmatrix}, \text{ where the blocks in this symmetric matrix are given by:}$$

$$\begin{aligned} \ddot{F}_{11, t} &= T^{-1} \mathbf{D}_{d, T}(\boldsymbol{\theta}_0)^{-1} \left(\text{diag}[\boldsymbol{\tau}] \text{diag}[\mathbf{d}_t(\boldsymbol{\theta})] (\ln t)^2 \right. \\ &\quad \left. - 2 \text{diag}[\tau_0] \text{diag}[\mathbf{d}_t(\boldsymbol{\theta})] \ln t \ln T \right) \mathbf{D}_{d, T}(\boldsymbol{\theta}_0)^{-1} \end{aligned}$$

$$\text{and } \ddot{F}_{12, t} = \ddot{F}_{21, t} = T^{-1} \mathbf{D}_{d, T}(\boldsymbol{\theta}_0)^{-1} \text{diag}[\mathbf{d}_t(\boldsymbol{\theta})] \ln t \mathbf{D}_{d, T}(\boldsymbol{\theta}_0)^{-1}.$$

$$(vi) \quad \text{Define a symmetric block matrix } \mathbf{M}_T = [\mathbf{M}_{T, ij}]_{1 \leq i, j \leq 3} = \mathbf{G}_{\gamma_0, T}^{-1} \left[\sum_{t=1}^T \dot{f}(\mathbf{w}_t, \gamma_0) \dot{f}(\mathbf{w}_t, \gamma_0)' \right] \mathbf{G}_{\gamma_0, T}^{-1} \text{ and a stacked vector}$$

$$\begin{aligned}
 z_T &= [z_{T,i}]_{1 \leq i \leq 3} = \mathbf{G}_{\gamma_0, T}'^{-1} \left[\sum_{t=1}^T \dot{f}(w_t, \gamma_0) u_t \right], \text{ where } M_{T,ij} = M_{T,ji}' \text{ and} \\
 M_{T,11} &= \frac{1}{T} \sum_{t=1}^T (\tau_0 \odot D_{d,T}(\theta_0)^{-1} d_t(\theta_0)) (\tau_0 \odot D_{d,T}(\theta_0)^{-1} d_t(\theta_0))' \left(\ln \frac{t}{T} \right)^2, \\
 M_{T,12} &= \frac{1}{T} \sum_{t=1}^T (\tau_0 \odot D_{d,T}(\theta_0)^{-1} d_t(\theta_0)) [D_{d,T}(\theta_0)^{-1} d_t(\theta_0)]' \ln \frac{t}{T}, \\
 M_{T,13} &= \frac{1}{T} \sum_{t=1}^T (\tau_0 \odot D_{d,T}(\theta_0)^{-1} d_t(\theta_0)) (D_{s,T}^{-1} s_t)' \ln \frac{t}{T}, \\
 M_{T,22} &= \frac{1}{T} \sum_{t=1}^T [D_{d,T}(\theta_0)^{-1} d_t(\theta_0)] [D_{d,T}(\theta_0)^{-1} d_t(\theta_0)]', \\
 M_{T,23} &= \frac{1}{T} \sum_{t=1}^T [D_{d,T}(\theta_0)^{-1} d_t(\theta_0)] (D_{s,T}^{-1} s_t)', \\
 M_{T,33} &= \frac{1}{T} \sum_{t=1}^T (D_{s,T}^{-1} s_t) (D_{s,T}^{-1} s_t)', \\
 z_{T,1} &= \frac{1}{\sqrt{T}} \sum_{t=1}^T (\tau_0 \odot D_{d,T}(\theta_0)^{-1} d_t(\theta_0)) \left(\ln \frac{t}{T} \right) u_t, \\
 z_{T,2} &= \frac{1}{\sqrt{T}} \sum_{t=1}^T [D_{d,T}(\theta_0)^{-1} d_t(\theta_0)] u_t, \\
 z_{T,3} &= \frac{1}{\sqrt{T}} \sum_{t=1}^T (D_{s,T}^{-1} s_t) u_t.
 \end{aligned}$$

Proof All results follow from linearity of the Hadamard product. ■

Lemma 4.4

For a constant $\delta > 0$, we define

$$\begin{aligned}
 \mathcal{N}_{\delta, T}(\gamma_0) = \{ \gamma \in \Gamma : & \|D_{d,T}(\theta_0)(\theta - \theta_0)\| < \delta T^{-1/2} \ln T, \|D_{s,T}(\phi - \phi_0)\| < \delta T^{-1/2} \ln T, \\
 & \|D_{d,T}(\theta_0)((\tau - \tau_0) + [\tau_0 \odot (\theta - \theta_0)] \ln T)\| < \delta T^{-1/2} \ln T \}, \quad (4.33)
 \end{aligned}$$

where $\gamma_0 \in \Gamma$ is fixed. In this lemma, we use \sup_{γ} to stand for the supremum over the set $\mathcal{N}_{\delta, T}(\gamma_0)$ for simplicity, i.e. $\sup_{\gamma \in \mathcal{N}_{\delta, T}(\gamma_0)}$. Under Assumption 4.2,

for all $k \geq 0$,

- (i) $\sup_{\gamma} T^{-1}(\ln T)^k \sum_{t=1}^T \left\| \mathbf{D}_{d,T}(\boldsymbol{\theta}_0)^{-1}(\mathbf{d}_t(\boldsymbol{\theta}) - \mathbf{d}_t(\boldsymbol{\theta}_0)) \right\|^2 = o(1),$
- (ii) $\sup_{\gamma} T^{-1}(\ln T)^k \sum_{t=1}^T \left\| \mathbf{D}_{d,T}(\boldsymbol{\theta}_0)^{-1}[\boldsymbol{\tau}_0 \odot (\mathbf{d}_t(\boldsymbol{\theta}) - \mathbf{d}_t(\boldsymbol{\theta}_0))] \ln \frac{t}{T} \right\|^2 = o(1),$
- (iii) $\sup_{\gamma} T^{-1}(\ln T)^k \sum_{t=1}^T \left\| \mathbf{D}_{d,T}(\boldsymbol{\theta}_0)^{-1}[(\boldsymbol{\tau} - \boldsymbol{\tau}_0) \odot \mathbf{d}_t(\boldsymbol{\theta}_0)] \ln t \right\|^2 = o(1),$
- (iv) $\sup_{\gamma} T^{-1}(\ln T)^k \sum_{t=1}^T \left\| \mathbf{D}_{d,T}(\boldsymbol{\theta}_0)^{-1}[(\boldsymbol{\tau} - \boldsymbol{\tau}_0) \odot (\mathbf{d}_t(\boldsymbol{\theta}) - \mathbf{d}_t(\boldsymbol{\theta}_0))] \ln t \right\|^2 = o(1),$
- (v) $\sup_{\gamma} \sum_{t=1}^T |f(\mathbf{w}_t, \gamma) - f(\mathbf{w}_t, \gamma_0)|^2 = O_p((\ln T)^4),$
- (vi) $\sup_{\gamma} T^{-2}(\ln T)^k \sum_{t=1}^T \left\| \mathbf{D}_{d,T}(\boldsymbol{\theta}_0)^{-2} \text{diag}[\mathbf{d}_t(\boldsymbol{\theta})] \right\|^2 = o(1),$
- (vii) $\sup_{\gamma} T^{-1}(\ln T) \left\| \sum_{t=1}^T \mathbf{D}_{d,T}(\boldsymbol{\theta}_0)^{-2} \text{diag}[\mathbf{d}_t(\boldsymbol{\theta})] u_t(\ln t)^k \right\| = o_p(1).$

Proof When needed, let $i \in \{1, 2, \dots, d\}$ be arbitrary. The i^{th} component of $\boldsymbol{\theta}_0$ and $\boldsymbol{\tau}_0$ are written θ_{0i} and τ_{0i} , respectively. (i) The i^{th} component of $\mathbf{D}_{d,T}(\boldsymbol{\theta}_0)^{-1}(\mathbf{d}_t(\boldsymbol{\theta}) - \mathbf{d}_t(\boldsymbol{\theta}_0))$ equals $T^{-\theta_{0i}}(t^{\theta_i} - t^{\theta_{0i}})$. By the mean-value theorem, we have

$$\begin{aligned}
 \frac{1}{T}(\ln T)^k \sum_{t=1}^T T^{-2\theta_{0i}} |t^{\theta_i} - t^{\theta_{0i}}|^2 &= \frac{1}{T}(\ln T)^k \sum_{t=1}^T \left(\frac{t}{T} \right)^{2\theta_{0i}} |t^{\theta_i - \theta_{0i}} - 1|^2 \\
 &\leq \frac{1}{T}(\ln T)^k \sum_{t=1}^T \left(\frac{t}{T} \right)^{2\theta_{0i}} (\ln t)^2 |\theta_i - \theta_{0i}|^2 \\
 &\leq \frac{1}{T}(\ln T)^{k+2} \sum_{t=1}^T \left(\frac{t}{T} \right)^{2\theta_{0i}} |\theta_i - \theta_{0i}|^2 \\
 &\leq (\ln T)^{k+2} \left(\sup_{\theta_L \leq \theta \leq \theta_U} \frac{1}{T} \sum_{t=1}^T \left(\frac{t}{T} \right)^{2\theta} \right) |\theta_i - \theta_{0i}|^2. \quad (4.34)
 \end{aligned}$$

The supremum in the RHS of (4.34) is bounded in view of Lemma 4.1(a). Moreover, if $\theta_i \in \mathcal{N}_{\delta,T}(\gamma_0)$, then $T^{\theta_{0i}} |\theta_i - \theta_{0i}| < \delta T^{-1/2} \ln T$ or equivalently $|\theta_i - \theta_{0i}|^2 < T^{-(1+2\theta_{0i})} \ln T$. We conclude that

$$\frac{1}{T}(\ln T)^k \sum_{t=1}^T T^{-2\theta_{0i}} |t^{\theta_i} - t^{\theta_{0i}}| \leq C(\ln T)^{k+3} T^{-(1+2\theta_{0i})} \quad (4.35)$$

which converges to zero because $1 + 2\theta_{0i} \geq 1 + 2\theta_L > 0$. The result follows since i was arbitrary. (ii) We again look at the i^{th} component of $D_{d,T}(\theta_0)^{-1}[\tau_0 \odot (d_i(\theta) - d_i(\theta_0))]\ln \frac{t}{T}$ and find

$$\begin{aligned} \frac{1}{T}(\ln T)^k \sum_{t=1}^T T^{-2\theta_{0i}} \tau_{0i}^2 |t^{\theta_i} - t^{\theta_{0i}}|^2 \left(\ln \frac{t}{T}\right)^2 \\ \leq (\ln T)^{k+2} \tau_{0i}^2 \left\{ \sup_{\theta_L \leq \theta \leq \theta_U} \frac{1}{T} \sum_{t=1}^T \left(\frac{t}{T}\right)^{2\theta} \left(\ln \frac{t}{T}\right)^2 \right\} |\theta_i - \theta_{0i}|^2 \end{aligned}$$

taking steps identical to those taken in (4.34). The supremum in the RHS is bounded since $\sup_{\theta_L \leq \theta \leq \theta_U} \left| \frac{1}{T} \sum_{t=1}^T \left(\frac{t}{T}\right)^{2\theta} \left(\ln \frac{t}{T}\right)^2 - \int_0^1 r^{2\theta} (\ln r)^2 dr \right| \rightarrow 0$ (a consequence of Lemma 4.1(iv)) and since $\int_0^1 r^{2\theta_{0i}} (\ln r)^2 dr$ is finite for all $\theta \in [\theta_L, \theta_U]$. The proof is easily completed after recalling that $|\theta_i - \theta_{0i}|^2 < T^{-(1+2\theta_{0i})} \ln T$ whenever $\theta_i \in \mathcal{N}_{\delta,T}(\gamma_0)$. (iii) The contribution of the i^{th} component of $D_{d,T}(\theta_0)^{-1}[(\tau - \tau_0) \odot d_i(\theta_0)]\ln t$ to the sum $\frac{1}{T}(\ln T)^k \sum_{t=1}^T \|D_{d,T}(\theta_0)^{-1}[(\tau - \tau_0) \odot d_i(\theta_0)]\ln t\|^2$ is

$$\frac{1}{T}(\ln T)^k \sum_{t=1}^T T^{-2\theta_{0i}} (\tau_i - \tau_{0i})^2 t^{2\theta_{0i}} (\ln t)^2 \leq (\ln T)^{k+2} \left\{ \sup_{\theta_L \leq \theta \leq \theta_U} \frac{1}{T} \sum_{t=1}^T \left(\frac{t}{T}\right)^{2\theta} \right\} (\tau_i - \tau_{0i})^2. \quad (4.36)$$

The supremum is bounded so it remains to say something about $(\tau_i - \tau_{0i})^2$. By the triangle inequality and the properties of norms (namely $\|\text{diag}[\mathbf{a}]\| \leq \|\mathbf{a}\|$), we have

$$\begin{aligned} \|D_{d,T}(\theta_0)(\tau - \tau_0)\| \\ \leq \|D_{d,T}(\theta_0)[\tau - \tau_0 + [\tau_0 \odot (\theta - \theta_0)]\ln T]\| + \|D_{d,T}(\theta_0) \text{diag}[\tau_0](\theta - \theta_0)\| \ln T \\ \leq \|D_{d,T}(\theta_0)[\tau - \tau_0 + [\tau_0 \odot (\theta - \theta_0)]\ln T]\| + \|\tau_0\| \|D_{d,T}(\theta_0)(\theta - \theta_0)\| \ln T \\ \leq \delta(1 + \|\tau_0\|)T^{-1/2}(\ln T)^2, \end{aligned} \quad (4.37)$$

for all $\tau \in \mathcal{N}_{\delta,T}(\gamma_0)$. (4.37) implies that $(\tau_i - \tau_{0i})^2 \leq \delta(1 + \|\tau_0\|)(\ln T)^4 T^{-(1+2\theta_{0i})}$ which goes to zero as $T \rightarrow \infty$. Now combine this finding with the RHS of (4.36) to establish the result. (iv) Use similar arguments as used in the proofs of (i) and (iii). (v) By definition of $f(w_t, \gamma)$ (see Lemma 4.3), it follows that

$$\begin{aligned} f(w_t, \gamma) - f(w_t, \gamma_0) &= d_t(\theta)' \tau + s_t' \phi - [d_t(\theta_0)' \tau_0 + s_t' \phi_0] \\ &= [d_t(\theta) - d_t(\theta_0)]'(\tau - \tau_0) + d_t(\theta_0)'(\tau - \tau_0) + [d_t(\theta) - d_t(\theta_0)]' \tau_0 + s_t'(\phi - \phi_0). \end{aligned}$$

and by the c_r -inequality that

$$\begin{aligned} \sum_{t=1}^T |f(\mathbf{w}_t, \gamma) - f(\mathbf{w}_t, \gamma_0)|^2 &\leq C \left\{ \sum_{t=1}^T |(\mathbf{d}_t(\boldsymbol{\theta}) - \mathbf{d}_t(\boldsymbol{\theta}_0))'(\boldsymbol{\tau} - \boldsymbol{\tau}_0)|^2 + \sum_{t=1}^T |\mathbf{d}_t(\boldsymbol{\theta}_0)'(\boldsymbol{\tau} - \boldsymbol{\tau}_0)|^2 \right. \\ &\quad \left. + \sum_{t=1}^T |(\mathbf{d}_t(\boldsymbol{\theta}) - \mathbf{d}_t(\boldsymbol{\theta}_0))' \boldsymbol{\tau}_0|^2 + \sum_{t=1}^T |\mathbf{s}'_t(\boldsymbol{\phi} - \boldsymbol{\phi}_0)|^2 \right\} =: C\{IVa + IVb + IVc + IVd\}. \end{aligned} \quad (4.38)$$

It remains to bound the terms IVa - IVd uniformly over $\mathcal{N}_{\delta,T}(\gamma_0)$. We repeatedly rely on $|\mathbf{a}'\mathbf{b}|^2 \leq \|\mathbf{a}\|^2 \|\mathbf{b}\|^2$ (Cauchy-Schwartz). We have

$$\begin{aligned} IVa &= \sum_{t=1}^T \left| [\mathbf{D}_{d,T}(\boldsymbol{\theta}_0)^{-1}(\mathbf{d}_t(\boldsymbol{\theta}) - \mathbf{d}_t(\boldsymbol{\theta}_0))]' [\mathbf{D}_{d,T}(\boldsymbol{\theta}_0)(\boldsymbol{\tau} - \boldsymbol{\tau}_0)] \right|^2 \\ &\leq T \left\| \mathbf{D}_{d,T}(\boldsymbol{\theta}_0)(\boldsymbol{\tau} - \boldsymbol{\tau}_0) \right\|^2 \frac{1}{T} \sum_{t=1}^T \left\| \mathbf{D}_{d,T}(\boldsymbol{\theta}_0)^{-1}(\mathbf{d}_t(\boldsymbol{\theta}) - \mathbf{d}_t(\boldsymbol{\theta}_0)) \right\|^2 \rightarrow 0 \end{aligned}$$

on $\mathcal{N}_{\delta,T}(\gamma_0)$ as $T \rightarrow \infty$ by (4.37) and Lemma 4.4(a). The bound on IVb is derived similarly, that is

$$\begin{aligned} IVb &\leq T \left\| \mathbf{D}_{d,T}(\boldsymbol{\theta}_0)(\boldsymbol{\tau} - \boldsymbol{\tau}_0) \right\|^2 \frac{1}{T} \sum_{t=1}^T \left\| \mathbf{D}_{d,T}(\boldsymbol{\theta}_0)^{-1} \mathbf{d}_t(\boldsymbol{\theta}_0) \right\|^2 \\ &= T \left\| \mathbf{D}_{d,T}(\boldsymbol{\theta}_0)(\boldsymbol{\tau} - \boldsymbol{\tau}_0) \right\|^2 \frac{1}{T} \sum_{t=1}^T \sum_{i=1}^d \left(\frac{t}{T} \right)^{2\theta_{0i}} \\ &\leq dT \left\| \mathbf{D}_{d,T}(\boldsymbol{\theta}_0)(\boldsymbol{\tau} - \boldsymbol{\tau}_0) \right\|^2 \frac{1}{T} \sum_{t=1}^T \left(\frac{t}{T} \right)^{2\theta_L}, \end{aligned}$$

which is $O((\ln T)^4)$ uniformly over $\mathcal{N}_{\delta,T}(\gamma_0)$ because $\left\| \mathbf{D}_{d,T}(\boldsymbol{\theta}_0)(\boldsymbol{\tau} - \boldsymbol{\tau}_0) \right\|^2 = O(T^{-1}(\ln T)^4)$ (using (4.37)) and $\frac{1}{T} \sum_{t=1}^T \left(\frac{t}{T} \right)^{2\theta_L}$ is bounded (see Lemma 4.1(i)). For

the third term, we establish

$$\begin{aligned}
 IVc &\leq \|\tau_0\|^2 \sum_{t=1}^T \sum_{i=1}^d |t^{\theta_i} - t^{\theta_{0i}}|^2 \\
 &= \|\tau_0\|^2 \sum_{i=1}^d \sum_{t=1}^T t^{2\theta_{0i}} |t^{\theta_i - \theta_{0i}} - 1|^2 \\
 &\leq \|\tau_0\|^2 \sum_{i=1}^d \sum_{t=1}^T t^{2\theta_{0i}} (\ln t)^2 (\theta_i - \theta_{0i})^2 \\
 &\leq \|\tau_0\|^2 (\ln T)^2 T \sum_{i=1}^d T^{2\theta_{0i}} (\theta_i - \theta_{0i})^2 \left[\frac{1}{T} \sum_{t=1}^T \left(\frac{t}{T} \right)^{2\theta_L} \right] \\
 &= \|\tau_0\|^2 (\ln T)^2 T \|D_{d,T}(\theta_0)(\theta - \theta_0)\|^2 \left[\frac{1}{T} \sum_{t=1}^T \left(\frac{t}{T} \right)^{2\theta_L} \right] = O((\ln T)^4)
 \end{aligned}$$

using the mean-value theorem, Lemma 4.1(i), and the fact that $\|D_{d,T}(\theta_0)(\theta - \theta_0)\|^2 = O(T^{-1}(\ln T)^2)$ on $\mathcal{N}_{\delta,T}(\gamma_0)$. Finally, the bound on IVd . We have

$$IVd \leq T \|D_{s,T}(\phi - \phi_0)\|^2 \frac{1}{T} \sum_{t=1}^T \|D_{s,T}^{-1} s_t\|^2 \leq T \|D_{s,T}(\phi - \phi_0)\|^2 \sum_{i=1}^m \sum_{j=1}^{p_i} \frac{1}{T} \sum_{t=1}^T \left(\frac{x_{it}}{\sqrt{T}} \right)^{2j},$$

which is $O_p((\ln T)^2)$ since $\|D_{s,T}(\phi - \phi_0)\|^2 \leq \delta T^{-1}(\ln T)^2$ on $\mathcal{N}_{\delta,T}(\gamma_0)$ and since Assumption 4.2 guarantees that all terms of the form $\frac{1}{T} \sum_{t=1}^T (x_{it}/\sqrt{T})^{2j}$ converge to integrals of Brownian motions. (vi) It follows by $\|\text{diag}[\mathbf{a}]\| \leq \|\mathbf{a}\|$, $\|\mathbf{AB}\| \leq \|\mathbf{A}\| \|\mathbf{B}\|$, the c_r -inequality, and the triangle inequality that

$$\begin{aligned}
 &\frac{1}{T^2} (\ln T)^k \sum_{t=1}^T \|D_{d,T}(\theta_0)^{-2} \text{diag}[\mathbf{d}_t(\theta)]\|^2 \leq \frac{1}{T^2} (\ln T)^k \sum_{t=1}^T \|D_{d,T}(\theta_0)^{-2} \mathbf{d}_t(\theta)\|^2 \\
 &\leq C \frac{1}{T} (\ln T)^k \|D_{d,T}(\theta_0)^{-1}\| \frac{1}{T} \sum_{t=1}^T \|D_{d,T}(\theta_0)^{-1} (\mathbf{d}_t(\theta) - \mathbf{d}_t(\theta_0))\|^2 \\
 &\quad + C \frac{1}{T} (\ln T)^k \|D_{d,T}(\theta_0)^{-1}\| \frac{1}{T} \sum_{t=1}^T \|D_{d,T}(\theta_0)^{-1} \mathbf{d}_t(\theta_0)\|^2
 \end{aligned} \tag{4.39}$$

Both terms are negligible because $\frac{1}{T} (\ln T)^k \|D_{d,T}(\theta_0)^{-1}\| \leq (\ln T)^k T^{-(1+\theta_L)} \rightarrow 0$ as $T \rightarrow \infty$, $\frac{1}{T} \sum_{t=1}^T \|D_{d,T}(\theta_0)^{-1} (\mathbf{d}_t(\theta) - \mathbf{d}_t(\theta_0))\|^2$ is $o(1)$ on $\mathcal{N}_{\delta,T}(\gamma_0)$ (Lemma 4.4(i)), and $\frac{1}{T} \sum_{t=1}^T \|D_{d,T}(\theta_0)^{-1} \mathbf{d}_t(\theta_0)\|^2$ is bounded in view of Lemma 4.1(i). (vii) Use similar

steps as in (4.39) to obtain

$$T^{-1}(\ln T) \left\| \sum_{t=1}^T \mathbf{D}_{d,T}(\boldsymbol{\theta}_0)^{-2} \text{diag}[\mathbf{d}_t(\boldsymbol{\theta})] u_t(\ln t)^k \right\| \leq (\ln T) T^{-(\theta_L + \frac{1}{2})} \times \\ \left\{ \underbrace{\left\| T^{-1/2} \sum_{t=1}^T \mathbf{D}_{d,T}(\boldsymbol{\theta}_0)^{-1} \mathbf{d}_t(\boldsymbol{\theta}_0) u_t(\ln t)^k \right\|}_{V(a)} + \underbrace{\left\| T^{-1/2} \sum_{t=1}^T \mathbf{D}_{d,T}(\boldsymbol{\theta}_0)^{-1} (\mathbf{d}_t(\boldsymbol{\theta}) - \mathbf{d}_t(\boldsymbol{\theta}_0)) u_t(\ln t)^k \right\|}_{V(b)} \right\}.$$

Note that the i^{th} component of the vector in $V(a)$ equals $T^{-1/2} \sum_{t=1}^T \left(\frac{t}{T}\right)^{\theta_{0i}} u_t(\ln t)^k$. Lemma 4.1(iii) and Chebyshev's inequality imply that $V(a)$ is $O_p((\ln T)^k)$. Furthermore, the mean-value theorem implies

$$\begin{aligned} \|V(b)\| &\leq \left\| [\mathbf{D}_{d,T}(\boldsymbol{\theta}_0)(\boldsymbol{\theta} - \boldsymbol{\theta}_0)] \odot \left[T^{-1/2} \sum_{t=1}^T \mathbf{D}_{d,T}(\boldsymbol{\theta}_0)^{-1} \mathbf{d}_t(\tilde{\boldsymbol{\theta}}) u_t(\ln t)^{k+1} \right] \right\| \\ &\leq \|\mathbf{D}_{d,T}(\boldsymbol{\theta}_0)(\boldsymbol{\theta} - \boldsymbol{\theta}_0)\| \left\| T^{-1/2} \sum_{t=1}^T \mathbf{D}_{d,T}(\boldsymbol{\theta}_0)^{-1} \mathbf{d}_t(\tilde{\boldsymbol{\theta}}) u_t(\ln t)^{k+1} \right\| \\ &\leq \|\mathbf{D}_{d,T}(\boldsymbol{\theta}_0)(\boldsymbol{\theta} - \boldsymbol{\theta}_0)\| \|\mathbf{D}_{d,T}(\boldsymbol{\theta}_0) \mathbf{D}_{d,T}(\tilde{\boldsymbol{\theta}})\| \left\| T^{-1/2} \sum_{t=1}^T \mathbf{D}_{d,T}(\boldsymbol{\theta}_0)^{-1} \mathbf{d}_t(\tilde{\boldsymbol{\theta}}) u_t(\ln t)^{k+1} \right\| \\ &\leq (\delta(\ln T) T^{-1/2}) T^{\|\boldsymbol{\theta} - \boldsymbol{\theta}_0\|_\infty} O_p((\ln T)^{k+2}) \end{aligned}$$

using Lemma 4.1(iii) and $\boldsymbol{\theta} \in \mathcal{N}_{\delta,T}(\gamma_0)$. Having established the stochastic orders of $V(a)$ and $V(b)$ it is straightforward to verify the claim in (vii). \blacksquare

Appendix 4.B Proof of Main Theorems

This section is devoted to the proof of Theorems 4.1 - 4.4. Before continuing, we recall the criterion function being defined as $Q_T(\gamma) = \frac{1}{2} \sum_{t=1}^T (y_t - f(\mathbf{w}_t, \gamma))^2$ with $f(\mathbf{w}_t, \gamma) = \mathbf{d}_t(\boldsymbol{\theta})' \boldsymbol{\tau} + \mathbf{s}_t' \boldsymbol{\phi}$ and $\mathbf{w}_t = [t, \mathbf{x}_t]'$. From theorem 3.1 in Chan and Wang (2015), we have $\mathbf{G}_{\gamma_0,T}(\widehat{\gamma}_T - \gamma_0) = \mathbf{M}_T^{-1} \mathbf{z}_T + o_p(1)$, if the following five conditions are fulfilled:¹⁸

- (i) $\|\mathbf{G}_{\gamma_0,T}^{-1}\| \rightarrow 0$ as $T \rightarrow \infty$;

¹⁸The original result in theorem 3.1 of Chan and Wang (2015) does not explicitly allow for deterministic trends and a scaling matrix $\mathbf{G}_{\gamma_0,T}$ that is parameter dependent. However, all steps in the proof remain valid after allowing for these features.

$$(ii) \sup_{\{\gamma: \|\mathbf{G}_{\gamma_0, T}(\gamma - \gamma_0)\| \leq k_T\}} \left\| \mathbf{G}_{\gamma_0, T}'^{-1} \sum_{t=1}^T [\dot{f}(\mathbf{w}_t, \gamma) \dot{f}(\mathbf{w}_t, \gamma)' - \dot{f}(\mathbf{w}_t, \gamma_0) \dot{f}(\mathbf{w}_t, \gamma_0)'] \mathbf{G}_{\gamma_0, T}^{-1} \right\| = o_p(1);$$

$$(iii) \sup_{\{\gamma: \|\mathbf{G}_{\gamma_0, T}(\gamma - \gamma_0)\| \leq k_T\}} \left\| \mathbf{G}_{\gamma_0, T}'^{-1} \sum_{t=1}^T \dot{f}(\mathbf{w}_t, \gamma) [f(\mathbf{w}_t, \gamma) - f(\mathbf{w}_t, \gamma_0)] \mathbf{G}_{\gamma_0, T}^{-1} \right\| = o_p(1);$$

$$(iv) \sup_{\{\gamma: \|\mathbf{G}_{\gamma_0, T}(\gamma - \gamma_0)\| \leq k_T\}} \left\| \mathbf{G}_{\gamma_0, T}'^{-1} \sum_{t=1}^T \dot{f}(\mathbf{w}_t, \gamma) u_t \mathbf{G}_{\gamma_0, T}^{-1} \right\| = o_p(1);$$

$$(v) \text{ for any } \alpha_i = [\alpha_{i1}, \dots, \alpha_{i, 2d+p}]' \in \mathbb{R}^{2d+p}, i = 1, 2, 3,$$

$$[\alpha_1' \mathbf{M}_T \alpha_2, \alpha_3' \mathbf{z}_T] \Rightarrow [\alpha_1' \mathbf{M} \alpha_2, \alpha_3' \mathbf{z}],$$

where $\mathbf{M} > 0$ a.s., $\mathbb{P}(z < \infty) = 1$ and

$$\begin{aligned} \mathbf{M}_T &= \mathbf{G}_{\gamma_0, T}'^{-1} \left[\sum_{t=1}^T \dot{f}(\mathbf{w}_t, \gamma_0) \dot{f}(\mathbf{w}_t, \gamma_0)' \right] \mathbf{G}_{\gamma_0, T}^{-1}, \\ \mathbf{z}_T &= \mathbf{G}_{\gamma_0, T}'^{-1} \left[\sum_{t=1}^T \dot{f}(\mathbf{w}_t, \gamma_0) u_t \right]. \end{aligned} \quad (4.40)$$

We make two remarks before verifying these conditions. First, we set $k_T = \delta \ln T$ and we will verify conditions (ii)-(iv) while replacing the supremum over the set

$$\tilde{\mathcal{N}}_{\delta, T}(\gamma_0) = \{\gamma \in \Gamma : \|\mathbf{G}_{\gamma_0, T}(\gamma - \gamma_0)\| \leq \delta \ln T\}$$

by a supremum over the set $\mathcal{N}_{\delta, T}(\gamma_0)$ given in (4.33). Since $\tilde{\mathcal{N}}_{\delta, T}(\gamma_0) \subset \mathcal{N}_{\delta, T}(\gamma_0)$, this replacement is innocuous. Second, note that $\|\mathbf{a} \odot \mathbf{b}\| \leq \|\mathbf{a} \odot \mathbf{b}\|_1 \leq \|\mathbf{a}\| \|\mathbf{b}\|$ holds for conformable vectors \mathbf{a} and \mathbf{b} .

Proof of Theorem 4.1 (i) From $\|\mathbf{L}_{\tau_0, T}\| \leq \|\mathbf{L}_{\tau_0, T}\|_{\mathcal{F}} = (2d + p + 2\|\tau_0\|^2(\ln T)^2)^{1/2} \leq C \ln T$, we have

$$\|\mathbf{G}_{\gamma_0, T}^{-1}\| \leq \|\mathbf{L}_{\tau_0, T}\| \|\mathbf{D}_{\theta_0, T}^{-1}\| \leq C(\ln T) T^{-1/2} (T^{-\theta_l} + T^{-1/2}) \xrightarrow{T \rightarrow \infty} 0.$$

(ii) Use $\sum_t \mathbf{a}_t \mathbf{a}_t' - \sum_t \mathbf{b}_t \mathbf{b}_t' = \sum_t (\mathbf{a}_t - \mathbf{b}_t)(\mathbf{a}_t - \mathbf{b}_t)' + \sum_t (\mathbf{a}_t - \mathbf{b}_t) \mathbf{b}_t' + \sum_t \mathbf{b}_t (\mathbf{a}_t - \mathbf{b}_t)'$ to obtain the bound

$$\begin{aligned}
& \left\| \mathbf{G}_{\gamma_0, T}^{-1} \sum_{t=1}^T \left[\dot{f}(\mathbf{w}_t, \gamma) \dot{f}(\mathbf{w}_t, \gamma)' - \dot{f}(\mathbf{w}_t, \gamma_0) \dot{f}(\mathbf{w}_t, \gamma_0)' \right] \mathbf{G}_{\gamma_0, T}^{-1} \right\| \\
& \leq \left\| \mathbf{D}_{\theta_0, T}^{-1} \left[\sum_{t=1}^T \mathbf{L}_{\tau_0, T} \left(\dot{f}(\mathbf{w}_t, \gamma) - \dot{f}(\mathbf{w}_t, \gamma_0) \right) \left(\dot{f}(\mathbf{w}_t, \gamma) - \dot{f}(\mathbf{w}_t, \gamma_0) \right)' \mathbf{L}_{\tau_0, T}' \right] \mathbf{D}_{\theta_0, T}^{-1} \right\| \\
& \quad + 2 \left\| \mathbf{D}_{\theta_0, T}^{-1} \left[\sum_{t=1}^T \mathbf{L}_{\tau_0, T} \left(\dot{f}(\mathbf{w}_t, \gamma) - \dot{f}(\mathbf{w}_t, \gamma_0) \right) \dot{f}(\mathbf{w}_t, \gamma_0)' \mathbf{L}_{\tau_0, T}' \right] \mathbf{D}_{\theta_0, T}^{-1} \right\| \\
& \leq \sum_{t=1}^T \left\| \mathbf{D}_{\theta_0, T}^{-1} \mathbf{L}_{\tau_0, T} \left(\dot{f}(\mathbf{w}_t, \gamma) - \dot{f}(\mathbf{w}_t, \gamma_0) \right) \right\|^2 \\
& \quad + 2 \left[(\ln T)^2 \sum_{t=1}^T \left\| \mathbf{D}_{\theta_0, T}^{-1} \mathbf{L}_{\tau_0, T} \left(\dot{f}(\mathbf{w}_t, \gamma) - \dot{f}(\mathbf{w}_t, \gamma_0) \right) \right\|^2 \right]^{1/2} \left[(\ln T)^{-2} \sum_{t=1}^T \left\| \mathbf{D}_{\theta_0, T}^{-1} \mathbf{L}_{\tau_0, T} \dot{f}(\mathbf{w}_t, \gamma_0) \right\|^2 \right]^{1/2}.
\end{aligned} \tag{4.41}$$

For any $k \geq 0$, the c_r -inequality and (4.32) yield a further upper bound as

$$\begin{aligned}
& (\ln T)^k \sum_{t=1}^T \left\| \mathbf{D}_{\theta_0, T}^{-1} \mathbf{L}_{\tau_0, T} \left(\dot{f}(\mathbf{w}_t, \gamma) - \dot{f}(\mathbf{w}_t, \gamma_0) \right) \right\|^2 \\
& \leq C (\ln T)^k \left\{ \frac{1}{T} \sum_{t=1}^T \left\| \mathbf{D}_{d, T}(\theta_0)^{-1} ((\boldsymbol{\tau} - \boldsymbol{\tau}_0) \odot (\mathbf{d}_t(\boldsymbol{\theta}) - \mathbf{d}_t(\theta_0))) \ln t \right\|^2 \right. \\
& \quad + \frac{1}{T} \sum_{t=1}^T \left\| \mathbf{D}_{d, T}(\theta_0)^{-1} (\mathbf{d}_t(\boldsymbol{\theta}) - \mathbf{d}_t(\theta_0)) \right\|^2 + \frac{1}{T} \sum_{t=1}^T \left\| \mathbf{D}_{d, T}(\theta_0)^{-1} (\boldsymbol{\tau}_0 \odot (\mathbf{d}_t(\boldsymbol{\theta}) - \mathbf{d}_t(\theta_0))) \ln \frac{t}{T} \right\|^2 \\
& \quad \left. + \frac{1}{T} \sum_{t=1}^T \left\| \mathbf{D}_{d, T}(\theta_0)^{-1} ((\boldsymbol{\tau} - \boldsymbol{\tau}_0) \odot \mathbf{d}_t(\theta_0)) \ln t \right\|^2 \right\}.
\end{aligned} \tag{4.42}$$

It follows from properties (i)-(iv) of Lemma 4.4 that each term in the RHS of (4.42) is $o(1)$ uniformly over $\mathcal{N}_{\delta, T}(\gamma_0)$. Moreover, since $\left(\ln \frac{t}{T} \right)^2 \leq 2(\ln T)^2$ for every $t = 1, 2, \dots, T$, we find that

$$\begin{aligned}
& (\ln T)^{-2} \sum_{t=1}^T \left\| \mathbf{D}_{\theta_0, T}^{-1} \mathbf{L}_{\tau_0, T} \dot{f}(\mathbf{w}_t, \gamma_0) \right\|^2 \\
& = (\ln T)^{-2} \frac{1}{T} \sum_{t=1}^T \left\| \mathbf{D}_{d, T}(\theta_0)^{-1} (\boldsymbol{\tau}_0 \odot \mathbf{d}_t(\theta_0)) \ln \frac{t}{T} \right\|^2 \\
& \quad + (\ln T)^{-2} \frac{1}{T} \sum_{t=1}^T \left\| \mathbf{D}_{d, T}(\theta_0)^{-1} \mathbf{d}_t(\theta_0) \right\|^2 + (\ln T)^{-2} \frac{1}{T} \sum_{t=1}^T \left\| \mathbf{D}_{s, T}^{-1} \mathbf{s}_t \right\|^2
\end{aligned} \tag{4.43}$$

is $O_p(1)$ in view of Lemmas 4.1(i) and 4.2(iii). The combination of (4.41), (4.42), and (4.43) leads to the desired result. **(iii)** The Cauchy-Schwarz inequality implies

$$\begin{aligned} & \left\| \mathbf{G}_{\gamma_0, T}^{\prime -1} \sum_{t=1}^T \ddot{f}(\mathbf{w}_t, \gamma) [f(\mathbf{w}_t, \gamma) - f(\mathbf{w}_t, \gamma_0)] \mathbf{G}_{\gamma_0, T}^{-1} \right\| \\ & \leq \left[(\ln T)^4 \sum_{t=1}^T \left\| \mathbf{G}_{\gamma_0, T}^{\prime -1} \ddot{f}(\mathbf{w}_t, \gamma) \mathbf{G}_{\gamma_0, T}^{-1} \right\|^2 \right]^{1/2} \left[(\ln T)^{-4} \sum_{t=1}^T |f(\mathbf{w}_t, \gamma) - f(\mathbf{w}_t, \gamma_0)|^2 \right]^{1/2}. \quad (4.44) \end{aligned}$$

By the identity in Lemma 4.3(v), we can bound the first term in the RHS of (4.44) as

$$\begin{aligned} (\ln T)^4 \sum_{t=1}^T \left\| \mathbf{G}_{\gamma_0, T}^{\prime -1} \ddot{f}(\mathbf{w}_t, \gamma) \mathbf{G}_{\gamma_0, T}^{-1} \right\|^2 & \leq C(\ln T)^4 \sum_{t=1}^T \|\ddot{\mathbf{F}}_{11,t}\|^2 + C(\ln T)^4 \sum_{t=1}^T \|\ddot{\mathbf{F}}_{12,t}\|^2 \\ & \leq C(\ln T)^8 \frac{1}{T^2} \sum_{t=1}^T \left\| \mathbf{D}_{d,T}(\boldsymbol{\theta}_0)^{-2} \text{diag}[\mathbf{d}_t(\boldsymbol{\theta})] \right\|^2, \quad (4.45) \end{aligned}$$

which is $o(1)$ uniformly over $\mathcal{N}_{\delta,T}(\gamma_0)$ by Lemma 4.4(vi). Note that the second inequality in (4.45) makes use of the facts that: (1) all matrices in $\ddot{\mathbf{F}}_{11,t}$ and $\ddot{\mathbf{F}}_{12,t}$ are diagonal and therefore commute, and (2) the triangle inequality gives

$$\begin{aligned} \|\text{diag}[\boldsymbol{\tau}]\| & \leq \|\boldsymbol{\tau}\| \\ & \leq \left\| \mathbf{D}_{d,T}(\boldsymbol{\theta}_0)^{-1} \right\| \left\| \mathbf{D}_{d,T}(\boldsymbol{\theta}_0)(\boldsymbol{\tau} - \boldsymbol{\tau}_0) \right\| + \|\boldsymbol{\tau}_0\| \leq C(\ln T)^2 T^{-(\theta_1+1/2)} + \|\boldsymbol{\tau}_0\| \leq C \end{aligned}$$

when $\gamma \in \mathcal{N}_{\delta,T}(\gamma_0)$ and T sufficiently large.¹⁹ The first term in (4.44) is $o(1)$ and the second is $O(1)$ (see Lemma 4.4(v)) over $\mathcal{N}_{\delta,T}(\gamma_0)$. The claim follows. **(iv)** By Lemma 4.3(v), we have

$$\left\| \mathbf{G}_{\gamma_0, T}^{\prime -1} \sum_{t=1}^T \ddot{f}(\mathbf{w}_t, \gamma) u_t \mathbf{G}_{\gamma_0, T}^{-1} \right\| \leq C \left(\left\| \sum_{t=1}^T \ddot{\mathbf{F}}_{11,t} u_t \right\| + \left\| \sum_{t=1}^T \ddot{\mathbf{F}}_{12,t} u_t \right\| \right)$$

¹⁹For illustration, we consider $(\ln T)^4 \sum_{t=1}^T \|\ddot{\mathbf{F}}_{11,t}\|^2$. By the definition of $\ddot{\mathbf{F}}_{11,t}$ in Lemma 4.3(v), we have $(\ln T)^4 \sum_{t=1}^T \|\ddot{\mathbf{F}}_{11,t}\|^2 \leq C \frac{(\ln T)^8}{T^2} \sum_{t=1}^T \left\| \mathbf{D}_{d,T}(\boldsymbol{\theta}_0)^{-1} \text{diag}[\boldsymbol{\tau}] \text{diag}[\mathbf{d}_t(\boldsymbol{\theta})] \mathbf{D}_{d,T}(\boldsymbol{\theta}_0)^{-1} \right\|^2 \leq C \frac{(\ln T)^8}{T^2} \|\text{diag}[\boldsymbol{\tau}]\|^2 \sum_{t=1}^T \left\| \mathbf{D}_{d,T}(\boldsymbol{\theta}_0)^{-1} \text{diag}[\mathbf{d}_t(\boldsymbol{\theta})] \mathbf{D}_{d,T}(\boldsymbol{\theta}_0)^{-1} \right\|^2$.

Using the definitions in Lemma 4.3(v) in a similar way as before, we obtain

$$\begin{aligned} \left\| \sum_{t=1}^T \ddot{F}_{11,t} u_t \right\| &\leq \|\text{diag}[\tau]\| \left\| T^{-1} \sum_{t=1}^T D_{d,T}(\theta_0)^{-2} \text{diag}[\mathbf{d}_t(\theta)] u_t (\ln t)^2 \right\| \\ &\quad + 2\|\text{diag}[\tau_0]\| (\ln T) \left\| T^{-1} \sum_{t=1}^T D_{d,T}(\theta_0)^{-2} \text{diag}[\mathbf{d}_t(\theta)] u_t \ln t \right\|, \end{aligned}$$

and $\left\| \sum_{t=1}^T \ddot{F}_{12,t} u_t \right\| = \left\| T^{-1} \sum_{t=1}^T D_{d,T}(\theta_0)^{-2} \text{diag}[\mathbf{d}_t(\theta)] u_t \ln t \right\|$. All relevant terms are $o_p(1)$ over $\mathcal{N}_{\delta,T}(\gamma_0)$ by Lemma 4.4(vii). (v) The convergence results in Lemma 4.2 applied to the definitions in Lemma 4.3(vi) provide $(M_T, z_T) \Rightarrow \left(\int_0^1 \dot{\mathbf{j}}(r; \gamma_0) \dot{\mathbf{j}}(r; \gamma_0)' dr, \int_0^1 \dot{\mathbf{j}}(r; \gamma_0) dB_u(r) + \mathcal{B}_{vu} \right)$ as $T \rightarrow \infty$. ■

Proof of Theorem 4.2 Changing the summation indices, we can express the one-sided long-run covariance estimator as

$$\widehat{\Delta}_T(\widehat{\gamma}_T, b_T) = \sum_{i=0}^{T-1} k\left(\frac{i}{b_T}\right) \left(\frac{1}{T} \sum_{t=1}^{T-i} \mathbf{V}_{t+i}(\widehat{\gamma}_T) \mathbf{V}_t(\widehat{\gamma}_T)' \right), \quad (4.46)$$

where we explicitly indicate the dependence on the parameter estimator $\widehat{\gamma}_T$ and bandwidth b_T . If we define $\widehat{\Sigma}_T(\widehat{\gamma}_T) = \frac{1}{T} \sum_{t=1}^T \mathbf{V}_t(\widehat{\gamma}_T) \mathbf{V}_t(\widehat{\gamma}_T)'$ and $\widehat{\Gamma}_T(\widehat{\gamma}_T, b_T) = \sum_{i=1}^{T-1} k\left(\frac{i}{b_T}\right) \left(\frac{1}{T} \sum_{t=1}^{T-i} \mathbf{V}_t(\widehat{\gamma}_T) \mathbf{V}_{t+i}(\widehat{\gamma}_T)' \right)$, then $\widehat{\Delta}_T(\widehat{\gamma}_T, b_T) = \widehat{\Sigma}_T(\widehat{\gamma}_T) + \widehat{\Gamma}_T(\widehat{\gamma}_T, b_T)$. We make two observations. First, we can write the two-sided estimator as

$$\widehat{\Omega}_T(\widehat{\gamma}_T, b_T) = \widehat{\Sigma}_T(\widehat{\gamma}_T) + \widehat{\Gamma}_T(\widehat{\gamma}_T, b_T) + \widehat{\Gamma}_T(\widehat{\gamma}_T, b_T)'. \quad (4.47)$$

It thus suffices to study the asymptotic behavior of $\widehat{\Sigma}_T(\widehat{\gamma}_T)$ and $\widehat{\Gamma}_T(\widehat{\gamma}_T, b_T)$ only. Second, the bottom right subblock of $\mathbf{V}_t(\gamma) \mathbf{V}_t(\gamma)'$ equals $\mathbf{v}_t \mathbf{v}_t'$ (no parameter estimation uncertainty here). The consistency results for this subblock are immediate from theorem 2 of Jansson (2002). We will therefore restrict our attention to $(1, 1)^{th}$ elements of $\widehat{\Sigma}_T(\widehat{\gamma}_T)$ and $\widehat{\Gamma}_T(\widehat{\gamma}_T, b_T)$. That is, we will show

$$\left[\widehat{\Sigma}_T(\widehat{\gamma}_T) \right]_{11} = \frac{1}{T} \sum_{t=1}^T \hat{u}_t^2 \rightarrow_p \mathbb{E}(u_t^2), \quad (4.48)$$

and

$$\left[\widehat{\Gamma}_T(\widehat{\gamma}_T, b_T) \right]_{11} = \sum_{i=1}^{T-1} k\left(\frac{i}{b_T}\right) \left(\frac{1}{T} \sum_{t=1}^{T-i} \hat{u}_{t+i} \hat{u}_t \right) \rightarrow_p \lim_{T \rightarrow \infty} \frac{1}{T} \sum_{i=1}^{T-1} \sum_{t=1}^{T-i} \mathbb{E}(u_{t+i} u_t). \quad (4.49)$$

The consistency proofs for the other elements in the first row/column of these matrices follows easily using similar arguments. The following result will be used throughout

$$\begin{aligned}
 \hat{u}_t - u_t &= z_t(\theta_0)' \begin{bmatrix} \tau_0 \\ \phi_0 \end{bmatrix} - z_t(\widehat{\theta}_T)' \begin{bmatrix} \widehat{\tau}_T \\ \widehat{\phi}_T \end{bmatrix} \\
 &= [z_t(\theta_0) - z_t(\widehat{\theta}_T)]' \begin{bmatrix} \tau_0 \\ \phi_0 \end{bmatrix} + [z_t(\theta_0) - z_t(\widehat{\theta}_T)]' \begin{bmatrix} \widehat{\tau}_T - \tau_0 \\ \widehat{\phi}_T - \phi_0 \end{bmatrix} - z_t(\theta_0)' \begin{bmatrix} \widehat{\tau}_T - \tau_0 \\ \widehat{\phi}_T - \phi_0 \end{bmatrix} \\
 &= [d_t(\theta_0) - d_t(\widehat{\theta}_T)]' \tau_0 + [d_t(\theta_0) - d_t(\widehat{\theta}_T)]' (\widehat{\tau}_T - \tau_0) - z_t(\theta_0)' \begin{bmatrix} \widehat{\tau}_T - \tau_0 \\ \widehat{\phi}_T - \phi_0 \end{bmatrix}. \quad (4.50)
 \end{aligned}$$

(i) We first show (4.48). Standard arguments provide $\frac{1}{T} \sum_{t=1}^T u_t^2 \rightarrow_p \mathbb{E}(u_t^2)$, so it suffices to show that $\frac{1}{T} \sum_{t=1}^T (u_t^2 - \hat{u}_t^2) = o_p(1)$. First, by Cauchy-Schwarz we find

$$\left| T^{-1} \sum_{t=1}^T (\hat{u}_t^2 - u_t^2) \right| \leq T^{-1} \sum_{t=1}^T (\hat{u}_t - u_t)^2 + 2 \left(T^{-1} \sum_{t=1}^T u_t^2 \right)^{1/2} \left(T^{-1} \sum_{t=1}^T (\hat{u}_t - u_t)^2 \right)^{1/2}.$$

It remains to establish $T^{-1} \sum_{t=1}^T (\hat{u}_t - u_t)^2 = o_p(1)$. Using (4.50), we see that the result follows if the following three statements are true:

$$\tau_0' \left(\frac{1}{T} \sum_{t=1}^T [d_t(\widehat{\theta}_T) - d_t(\theta_0)][d_t(\widehat{\theta}_T) - d_t(\theta_0)]' \right) \tau_0 = o_p(1), \quad (4.51a)$$

$$(\widehat{\tau}_T - \tau_0)' \left(\frac{1}{T} \sum_{t=1}^T [d_t(\widehat{\theta}_T) - d_t(\theta_0)][d_t(\theta_0) - d_t(\widehat{\theta}_T)]' \right) (\widehat{\tau}_T - \tau_0) = o_p(1), \quad (4.51b)$$

$$\begin{bmatrix} \widehat{\tau}_T - \tau_0 \\ \widehat{\phi}_T - \phi_0 \end{bmatrix}' \frac{1}{T} \sum_{t=1}^T z_t(\theta_0) z_t(\theta_0)' \begin{bmatrix} \widehat{\tau}_T - \tau_0 \\ \widehat{\phi}_T - \phi_0 \end{bmatrix} = o_p(1). \quad (4.51c)$$

We first look at the norm of the $(i, j)^{th}$ component of $\frac{1}{T} \sum_{t=1}^T [d_t(\widehat{\theta}_T) - d_t(\theta_0)][d_t(\widehat{\theta}_T) - d_t(\theta_0)]'$, i.e.

$$\begin{aligned}
 \left| \frac{1}{T} \sum_{t=1}^T (t^{\widehat{\theta}_i} - t^{\theta_{0i}})(t^{\widehat{\theta}_j} - t^{\theta_{0j}}) \right| &= \frac{1}{T} \sum_{t=1}^T t^{\theta_{0i} + \theta_{0j}} |t^{\widehat{\theta}_i - \theta_{0i}} - 1| |t^{\widehat{\theta}_j - \theta_{0j}} - 1| \\
 &\leq C \frac{(\ln T)^2}{T} |T^{\theta_{0i} + \frac{1}{2}}(\widehat{\theta}_i - \theta_{0i})| |T^{\theta_{0j} + \frac{1}{2}}(\widehat{\theta}_j - \theta_{0j})| \frac{1}{T} \sum_{t=1}^T \left(\frac{t}{T} \right)^{\theta_{0i} + \theta_{0j}}. \quad (4.52)
 \end{aligned}$$

By the convergence result from Theorem 4.1 and Lemma 4.1(i), (4.52) is $O_p(T^{-1}(\ln T)^2)$. The statements in (4.51a) and (4.51b) follow easily. We subsequently introduce the scaling matrix $\mathcal{D}_{\theta_0, T} = \begin{bmatrix} D_{d, T}(\theta_0) & \mathbf{0} \\ \mathbf{0} & D_{s, T} \end{bmatrix}$. The LHS of (4.51c)

can now be expressed as

$$\left(\mathcal{D}_{\theta_0, T} \begin{bmatrix} \widehat{\tau}_T - \tau_0 \\ \widehat{\phi}_T - \phi_0 \end{bmatrix} \right)' \left(T^{-1} \sum_{t=1}^T [\mathcal{D}_{\theta_0, T}^{-1} z_t(\theta_0)] [\mathcal{D}_{\theta_0, T}^{-1} z_t(\theta_0)]' \right) \left(\mathcal{D}_{\theta_0, T} \begin{bmatrix} \widehat{\tau}_T - \tau_0 \\ \widehat{\phi}_T - \phi_0 \end{bmatrix} \right). \quad (4.53)$$

The results in Lemma 4.2(iii) imply that $T^{-1} \sum_{t=1}^T [\mathcal{D}_{\theta_0, T}^{-1} z_t(\theta_0)] [\mathcal{D}_{\theta_0, T}^{-1} z_t(\theta_0)]' \Rightarrow \int_0^1 \widetilde{\mathbf{j}}(r; \theta_0) \widetilde{\mathbf{j}}(r; \theta_0)' dr$, where $\widetilde{\mathbf{j}}(r; \theta_0) = [\mathbf{d}(r; \theta_0)', \mathbf{B}'_{(1)}(r), \dots, \mathbf{B}'_{(m)}(r)]'$. A comparison of the elements of $\mathcal{D}_{\theta_0, T} \begin{bmatrix} \widehat{\tau}_T - \tau_0 \\ \widehat{\phi}_T - \phi_0 \end{bmatrix}$ with the convergence rates of these estimators leads us to conclude that (4.51c) is also true.

(ii) To prove (4.49), we again show that the parameter estimation error is asymptotically negligible. If this holds, then the remainder of the proof follows from Jansson (2002). This said, we write $\sum_{i=1}^{T-1} k\left(\frac{i}{b_T}\right) \left(\frac{1}{T} \sum_{t=1}^{T-i} (\hat{u}_{t+i} \hat{u}_t - u_{t+i} u_t) \right) = I + II + III$, where

$$I = \sum_{i=1}^{T-1} k\left(\frac{i}{b_T}\right) \left(\frac{1}{T} \sum_{t=1}^{T-i} u_{t+i} (\hat{u}_t - u_t) \right), \quad (4.54)$$

$$II = \sum_{i=1}^{T-1} k\left(\frac{i}{b_T}\right) \left(\frac{1}{T} \sum_{t=1}^{T-i} (\hat{u}_{t+i} - u_{t+i}) u_t \right), \quad (4.55)$$

$$III = \sum_{i=1}^{T-1} k\left(\frac{i}{b_T}\right) \left(\frac{1}{T} \sum_{t=1}^{T-i} (\hat{u}_{t+i} - u_{t+i}) (\hat{u}_t - u_t) \right). \quad (4.56)$$

We provide details for $I = o_p(1)$ and omit the explicit proofs for II and III . Similar (and tedious) calculations are applicable there. Using (4.50), we can further decompose $I = I_a + I_b - I_c$ with

$$\begin{aligned} I_a &= \sum_{i=1}^{T-1} k\left(\frac{i}{b_T}\right) \left(\frac{1}{T} \sum_{t=1}^{T-i} u_{t+i} [\mathbf{d}_t(\theta_0) - \mathbf{d}_t(\widehat{\theta}_T)]' \tau_0 \right), \\ I_b &= \sum_{i=1}^{T-1} k\left(\frac{i}{b_T}\right) \left(\frac{1}{T} \sum_{t=1}^{T-i} u_{t+i} [\mathbf{d}_t(\theta_0) - \mathbf{d}_t(\widehat{\theta}_T)]' (\widehat{\tau}_T - \tau_0) \right), \\ I_c &= \sum_{i=1}^{T-1} k\left(\frac{i}{b_T}\right) \left(\frac{1}{T} \sum_{t=1}^{T-i} u_{t+i} z_t(\theta_0)' \begin{bmatrix} \widehat{\tau}_T - \tau_0 \\ \widehat{\phi}_T - \phi_0 \end{bmatrix} \right). \end{aligned}$$

We adjust Hansen's (1992) argument slightly and look at the quantities $\frac{T^{1/2}}{b_T \ln T} |I_i|$ for $i \in \{a, b, c\}$. If these quantities are stochastically bounded, then the result follows

because $\frac{b_T \ln T}{T^{1/2}} \rightarrow 0$ by Assumption 4.3. By the Cauchy-Schwartz inequality, we have

$$\begin{aligned} \frac{T^{1/2}}{b_T \ln T} |I_a| &\leq \frac{T^{1/2}}{b_T \ln T} \sum_{i=1}^{T-1} \left| k\left(\frac{i}{b_T}\right) \right| \left| T^{-1} \sum_{t=1}^{T-i} u_{t+i} [\mathbf{d}_t(\boldsymbol{\theta}_0) - \mathbf{d}_t(\widehat{\boldsymbol{\theta}}_T)]' \boldsymbol{\tau}_0 \right| \\ &\leq \frac{T^{1/2}}{b_T \ln T} \sum_{i=1}^{T-1} \left| k\left(\frac{i}{b_T}\right) \right| \left(T^{-1} \sum_{t=1}^{T-i} u_{t+i}^2 \right)^{1/2} \left[T^{-1} \sum_{t=1}^{T-i} ([\mathbf{d}_t(\boldsymbol{\theta}_0) - \mathbf{d}_t(\widehat{\boldsymbol{\theta}}_T)]' \boldsymbol{\tau}_0)^2 \right]^{1/2} \\ &\leq \frac{1}{b_T} \sum_{i=1}^{T-1} \left| k\left(\frac{i}{b_T}\right) \right| \left(T^{-1} \sum_{t=1}^T u_t^2 \right)^{1/2} \left\{ \boldsymbol{\tau}_0' \left[(\ln T)^{-2} \sum_{t=1}^T [\mathbf{d}_t(\widehat{\boldsymbol{\theta}}_T) - \mathbf{d}_t(\boldsymbol{\theta}_0)][\mathbf{d}_t(\widehat{\boldsymbol{\theta}}_T) - \mathbf{d}_t(\boldsymbol{\theta}_0)]' \right] \boldsymbol{\tau}_0 \right\}^{1/2}. \end{aligned}$$

The upper bound above is bounded in probability due to lemma 1 of Jansson (2002), the fact that $\frac{1}{T} \sum_{t=1}^T u_t^2 \rightarrow_p \mathbb{E}(u_t^2)$, and (4.52). Similarly, use

$$\begin{aligned} \frac{T^{1/2}}{b_T \ln T} |I_b| &\leq \frac{1}{b_T} \sum_{i=1}^{T-1} \left| k\left(\frac{i}{b_T}\right) \right| \left(T^{-1} \sum_{t=1}^T u_t^2 \right)^{1/2} \times \\ &\quad \left\{ (\widehat{\boldsymbol{\tau}}_T - \boldsymbol{\tau}_0)' \left[(\ln T)^{-2} \sum_{t=1}^T [\mathbf{d}_t(\widehat{\boldsymbol{\theta}}_T) - \mathbf{d}_t(\boldsymbol{\theta}_0)][\mathbf{d}_t(\widehat{\boldsymbol{\theta}}_T) - \mathbf{d}_t(\boldsymbol{\theta}_0)]' \right] (\widehat{\boldsymbol{\tau}}_T - \boldsymbol{\tau}_0) \right\}^{1/2}. \end{aligned}$$

to show that $\frac{T^{1/2}}{b_T \ln T} |I_b| = O_p(1)$. Finally, we have

$$\begin{aligned} \frac{T^{1/2}}{b_T \ln T} |I_c| &\leq \frac{T^{1/2}}{b_T \ln T} \sum_{i=1}^{T-1} \left| k\left(\frac{i}{b_T}\right) \right| \left| \frac{1}{T} \sum_{t=1}^{T-i} u_{t+i} \mathbf{z}_t(\boldsymbol{\theta}_0)' \begin{bmatrix} \widehat{\boldsymbol{\tau}}_T - \boldsymbol{\tau}_0 \\ \widehat{\boldsymbol{\phi}}_T - \boldsymbol{\phi}_0 \end{bmatrix} \right| \\ &\leq \frac{1}{b_T} \sum_{i=1}^{T-1} \left| k\left(\frac{i}{b_T}\right) \right| \left(T^{-1} \sum_{t=1}^T u_t^2 \right)^{1/2} \times \\ &\quad \left\{ \left(\frac{T^{1/2}}{\ln T} \mathcal{D}_{\boldsymbol{\theta}_0, T} \begin{bmatrix} \widehat{\boldsymbol{\tau}}_T - \boldsymbol{\tau}_0 \\ \widehat{\boldsymbol{\phi}}_T - \boldsymbol{\phi}_0 \end{bmatrix} \right)' \left(T^{-1} \sum_{t=1}^T [\mathcal{D}_{\boldsymbol{\theta}_0, T}^{-1} \mathbf{z}_t(\boldsymbol{\theta}_0)][\mathcal{D}_{\boldsymbol{\theta}_0, T}^{-1} \mathbf{z}_t(\boldsymbol{\theta}_0)]' \right) \left(\frac{T^{1/2}}{\ln T} \mathcal{D}_{\boldsymbol{\theta}_0, T} \begin{bmatrix} \widehat{\boldsymbol{\tau}}_T - \boldsymbol{\tau}_0 \\ \widehat{\boldsymbol{\phi}}_T - \boldsymbol{\phi}_0 \end{bmatrix} \right) \right\}^{1/2}. \end{aligned}$$

Now note that $\frac{T^{1/2}}{\ln T} \mathbf{D}_{d,T}(\boldsymbol{\theta}_0)(\widehat{\boldsymbol{\tau}}_T - \boldsymbol{\tau}_0)$ and $T^{1/2} \mathbf{D}_{s,T}(\widehat{\boldsymbol{\phi}}_T - \boldsymbol{\phi}_0)$ are $O_p(1)$. This completes the proof. \blacksquare

Proof of Theorem 4.3 For brevity, we define

$$\begin{aligned} &\widehat{\mathcal{J}}_N(\widehat{\boldsymbol{\gamma}}_T, \widehat{\boldsymbol{\Omega}}, \widehat{\boldsymbol{\Delta}}_{vu}^-) \\ &= \left\{ \mathbf{G}'_{\widehat{\boldsymbol{\gamma}}, N} \left[\sum_{n=1}^N \dot{f}(\widehat{\mathbf{w}}_n, \widehat{\boldsymbol{\gamma}}_T) \dot{f}(\widehat{\mathbf{w}}_n, \widehat{\boldsymbol{\gamma}}_T)' \right] \mathbf{G}_{\widehat{\boldsymbol{\gamma}}, N}^{-1} \right\}^{-1} \left\{ \mathbf{G}_{\widehat{\boldsymbol{\gamma}}, N}^{\prime -1} \left[\sum_{n=1}^N \dot{f}(\widehat{\mathbf{w}}_n, \widehat{\boldsymbol{\gamma}}_T) \widehat{\boldsymbol{\mu}}_n \right] + \widehat{\mathcal{B}}_{vu}^- \right\}. \end{aligned}$$

We first show $\widehat{\mathcal{J}}_N(\widehat{\gamma}_T, \widehat{\Omega}, \widehat{\Delta}_{vu}^-) = \widehat{\mathcal{J}}_N(\gamma_0, \widehat{\Omega}, \widehat{\Delta}_{vu}^-) + o_p^*(1)$. Direct calculation yields

$$\begin{aligned} \mathbf{R}_N &:= \mathbf{G}_{\gamma_0, N} \mathbf{G}_{\widehat{\gamma}, N}^{-1} \\ &= \begin{bmatrix} \mathbf{D}_{d, N}(\theta_0) & & \\ \mathbf{D}_{d, N}(\theta_0) \text{diag}[\tau_0] \ln N & \mathbf{D}_{d, N}(\theta_0) & \\ \mathbf{O}_{p \times d} & \mathbf{O}_{p \times d} & \mathbf{D}_{s, N} \end{bmatrix} \\ &\quad \times \begin{bmatrix} \mathbf{D}_{d, N}(\widehat{\theta}_T)^{-1} & & \\ -\mathbf{D}_{d, N}(\widehat{\theta}_T)^{-1} \text{diag}[\widehat{\tau}_T] \ln N & \mathbf{D}_{d, N}(\widehat{\theta}_T)^{-1} & \\ \mathbf{O}_{p \times d} & \mathbf{O}_{p \times d} & \mathbf{D}_{s, N}^{-1} \end{bmatrix} \\ &= \begin{bmatrix} \mathbf{D}_{d, N}(\theta_0) \mathbf{D}_{d, N}(\widehat{\theta}_T)^{-1} & & \\ \mathbf{D}_{d, N}(\theta_0) \mathbf{D}_{d, N}(\widehat{\theta}_T)^{-1} \text{diag}[\tau_0 - \widehat{\tau}_T] \ln N & \mathbf{D}_{d, N}(\theta_0) \mathbf{D}_{d, N}(\widehat{\theta}_T)^{-1} & \\ \mathbf{O}_{p \times d} & \mathbf{O}_{p \times d} & \mathbf{I}_p \end{bmatrix}. \end{aligned}$$

We have $\mathbf{R}_N \rightarrow_p \mathbf{I}_{2d+p}$. To see this, note that (1) a typical diagonal element of $\mathbf{D}_{d, N}(\theta_0) \mathbf{D}_{d, N}(\widehat{\theta}_T)^{-1}$ is $N^{\theta_{0i} - \widehat{\theta}_i}$ for which $N^{|\widehat{\theta}_i - \theta_{0i}|} = \exp(|\widehat{\theta}_i - \theta_{0i}| \ln N) \leq \exp((\ln N)^{-(\theta_L + \frac{1}{2})} (\frac{N}{T})^{\theta_L + \frac{1}{2}} |T^{\theta_{0i} + \frac{1}{2}} (\widehat{\theta}_i - \theta_{0i})|) \rightarrow_p 1$, and (2) $\|\text{diag}[\tau_0 - \widehat{\tau}_T] \ln N\| \leq \|\mathbf{D}_{d, T}(\theta_0)^{-1}\| \frac{\ln T}{T^{1/2}} (\ln N) \left\| \frac{T^{1/2}}{\ln T} \mathbf{D}_{d, T}(\theta_0) (\widehat{\tau}_T - \tau_0) \right\| \leq C \frac{\ln T}{T^{1/2 + \theta_L}} (\ln N) O_p(1) = o_p(1)$. Define $\widetilde{\mathcal{N}}_{\delta^*, N}(\gamma_0)$ similarly to $\widetilde{\mathcal{N}}_{\delta, T}(\gamma_0)$ (page 198 below (4.40)). Consequently, if there exists a constant $\delta^* > 0$ such that $\mathbb{P}(\widehat{\gamma}_T \in \widetilde{\mathcal{N}}_{\delta^*, N}(\gamma_0)) \rightarrow 1$, then

$$\begin{aligned} \mathbf{G}_{\widehat{\gamma}, N}^{\prime -1} \left[\sum_{n=1}^N \dot{f}(\widehat{\mathbf{w}}_n, \widehat{\gamma}_T) \dot{f}(\widehat{\mathbf{w}}_n, \widehat{\gamma}_T)' \right] \mathbf{G}_{\widehat{\gamma}, N}^{-1} &= \mathbf{R}_N' \mathbf{G}_{\gamma_0, N}^{\prime -1} \left[\sum_{n=1}^N \dot{f}(\widehat{\mathbf{w}}_n, \widehat{\gamma}_T) \dot{f}(\widehat{\mathbf{w}}_n, \widehat{\gamma}_T)' \right] \mathbf{G}_{\gamma_0, N}^{-1} \mathbf{R}_N \\ &= \mathbf{R}_N' \left\{ \mathbf{G}_{\gamma_0, N}^{\prime -1} \left[\sum_{n=1}^N \dot{f}(\widehat{\mathbf{w}}_n, \gamma_0) \dot{f}(\widehat{\mathbf{w}}_n, \gamma_0)' \right] \mathbf{G}_{\gamma_0, N}^{-1} + o_p(1) \right\} \mathbf{R}_N \\ &= \mathbf{G}_{\gamma_0, N}^{\prime -1} \left[\sum_{n=1}^N \dot{f}(\widehat{\mathbf{w}}_n, \gamma_0) \dot{f}(\widehat{\mathbf{w}}_n, \gamma_0)' \right] \mathbf{G}_{\gamma_0, N}^{-1} + o_p(1), \end{aligned} \quad (4.57)$$

where the second equality follows from the same arguments in (4.41). The condition $\lim_{T \rightarrow \infty} \mathbb{P}(\widehat{\gamma}_T \in \widetilde{\mathcal{N}}_{\delta^*, N}(\gamma_0)) = 1$ is easily satisfied because $\|\mathbf{G}_{\gamma_0, N}(\widehat{\gamma}_T - \gamma_0)\| \leq \|\mathbf{G}_{\gamma_0, N} \mathbf{G}_{\gamma_0, T}^{-1}\| \|\mathbf{G}_{\gamma_0, T}(\widehat{\gamma}_T - \gamma_0)\| \leq C \delta \ln T =: \delta^* \ln T$, where

$$\begin{aligned} \mathbf{G}_{\gamma_0, N} \mathbf{G}_{\gamma_0, T}^{-1} &= (N/T)^{1/2} \\ &\quad \times \begin{bmatrix} \mathbf{D}_{d, N}(\theta_0) \mathbf{D}_{d, T}(\theta_0)^{-1} & & \\ \mathbf{D}_{d, N}(\theta_0) \mathbf{D}_{d, T}(\theta_0)^{-1} \text{diag}[\tau_0] (\ln(N/T)) & \mathbf{D}_{d, N}(\theta_0) \mathbf{D}_{d, T}(\theta_0)^{-1} & \\ \mathbf{O}_{p \times d} & \mathbf{O}_{p \times d} & \mathbf{D}_{s, N} \mathbf{D}_{s, T}^{-1} \end{bmatrix}, \end{aligned}$$

and thus, by the norm property $\|\cdot\|^2 \leq \|\cdot\|_{\mathcal{F}}^2$,

$$\begin{aligned} \|G_{\gamma_0, N} G_{\gamma_0, T}^{-1}\|^2 &\leq C(N/T) \left(\|D_{d, N}(\theta_0) D_{d, T}^{-1}(\theta_0)\|_{\mathcal{F}}^2 (\ln(N/T))^2 + \|D_{s, N} D_{s, T}^{-1}\|_{\mathcal{F}}^2 \right) \\ &\leq C(N/T)^{1+2\theta_L} (\ln(N/T))^2 + (N/T)^2 \leq C. \end{aligned}$$

Next, we consider $G_{\widehat{\gamma}, N}^{\prime-1} \sum_{n=1}^N \dot{f}(\widehat{w}_n, \widehat{\gamma}_T) \widehat{\mu}_n$, or equivalently

$$\begin{aligned} G_{\widehat{\gamma}, N}^{\prime-1} \left[\sum_{n=1}^N \dot{f}(\widehat{w}_n, \widehat{\gamma}_T) \widehat{\mu}_n \right] \\ = R_N' \left[G_{\gamma_0, N}^{\prime-1} \sum_{n=1}^N \dot{f}(\widehat{w}_n, \gamma_0) \widehat{\mu}_n + G_{\gamma_0, N}^{\prime-1} \sum_{n=1}^N (\dot{f}(\widehat{w}_n, \widehat{\gamma}_T) - \dot{f}(\widehat{w}_n, \gamma_0)) \widehat{\mu}_n \right]. \end{aligned}$$

We know $R_N \rightarrow_p I_{2d+p}$. Moreover, by the triangle and Cauchy-Schwartz inequalities,

$$\begin{aligned} \left\| G_{\gamma_0, N}^{\prime-1} \sum_{n=1}^N (\dot{f}(\widehat{w}_n, \widehat{\gamma}_T) - \dot{f}(\widehat{w}_n, \gamma_0)) \widehat{\mu}_n \right\| \\ \leq \left[N \sum_{n=1}^N \left\| G_{\gamma_0, N}^{\prime-1} (\dot{f}(\widehat{w}_n, \widehat{\gamma}_T) - \dot{f}(\widehat{w}_n, \gamma_0)) \right\|^2 \right]^{1/2} \left(N^{-1} \sum_{n=1}^N \widehat{\mu}_n^2 \right)^{1/2}. \quad (4.58) \end{aligned}$$

Using $G_{\gamma_0, N} = D_{\theta_0, N} L_{\tau_0, N}^{\prime-1}$, we see from (4.32) that the first term in the RHS of (4.58) does not depend on $\{e_1, \dots, e_N\}$. As in (4.42), we also have $N \sum_{n=1}^N \left\| G_{\gamma_0, N}^{\prime-1} (\dot{f}(\widehat{w}_n, \widehat{\gamma}_T) - \dot{f}(\widehat{w}_n, \gamma_0)) \right\|^2 \leq C \sum_{i=1}^4 I_i$ with

$$\begin{aligned} I_1 &= \sum_{n=1}^N \left\| D_{d, N}(\theta_0)^{-1} ((\widehat{\tau}_T - \tau_0) \odot (d_n(\widehat{\theta}_T) - d_n(\theta_0))) \ln n \right\|^2 = O_p \left(\frac{(\ln T)^6}{T^{1+2\theta_L}} \frac{N}{T^{1+2\theta_L}} \right) = o_p(1), \\ I_2 &= \sum_{n=1}^N \left\| D_{d, N}(\theta_0)^{-1} (d_n(\widehat{\theta}_T) - d_n(\theta_0)) \right\|^2 = O_p \left((\ln T)^2 \frac{N}{T^{1+2\theta_L}} \right) = o_p(1), \\ I_3 &= \sum_{n=1}^N \left\| D_{d, N}(\theta_0)^{-1} (\tau_0 \odot (d_n(\widehat{\theta}_T) - d_n(\theta_0))) \ln \frac{n}{N} \right\|^2 = O_p \left((\ln T)^4 \frac{N}{T^{1+2\theta_L}} \right) = o_p(1), \\ I_4 &= \sum_{n=1}^N \left\| D_{d, N}(\theta_0)^{-1} ((\widehat{\tau}_T - \tau_0) \odot d_n(\theta_0)) \ln n \right\|^2 = O_p \left((\ln T)^4 \frac{N}{T^{1+2\theta_L}} \right) = o_p(1), \end{aligned}$$

and where stochastic orders are established as in Lemma 4.4 (i) to (iv) (and thus omitted). For the second term in the RHS of (4.58), note that

$\frac{1}{N} \sum_{n=1}^N \widehat{\mu}_n^2 \leq \frac{1}{N} \sum_{n=1}^N (\widehat{\mu}_n^2 + \widehat{v}_n' \widehat{v}_n) = \frac{1}{N} \sum_{n=1}^N \mathbf{e}_n' \widehat{\Omega}_T \mathbf{e}_n \leq \|\widehat{\Omega}_T\| \left(\frac{1}{N} \sum_{n=1}^N \mathbf{e}_n' \mathbf{e}_n \right) = O_p^*(1)$ because $\widehat{\Omega}_T \rightarrow_p \Omega$ and $\mathbf{e}_n \stackrel{i.i.d.}{\sim} N(\mathbf{0}, \mathbf{I}_{m+1})$. Overall, we have $G_{\widehat{\gamma}, N}^{-1} \left[\sum_{n=1}^N \dot{f}(\widehat{w}_n, \widehat{\gamma}_T) \widehat{\mu}_n \right] = G_{\gamma_0, N}^{-1} \sum_{n=1}^N \dot{f}(\widehat{w}_n, \gamma_0) \widehat{\mu}_n + o_p^*(1)$. Combining this result with (4.57) gives $\widehat{\mathcal{J}}_N(\widehat{\gamma}_T, \widehat{\Omega}, \widehat{\Delta}_{vu}^-) = \widehat{\mathcal{J}}_N(\gamma_0, \widehat{\Omega}, \widehat{\Delta}_{vu}^-) + o_p^*(1)$.

Finally, we consider $\widehat{\mathcal{J}}_N(\gamma_0, \widehat{\Omega}, \widehat{\Delta}_{vu}^-)$ itself. By independence between $\{\mathbf{e}_n\}$ and $\{\widehat{\Omega}, \widehat{\Delta}_{vu}^-\}$, consistency of $\widehat{\Omega}$, and a FCLT for the i.i.d. sequence, we have

$$\frac{1}{\sqrt{N}} \sum_{n=1}^{[rN]} \begin{bmatrix} \widehat{\mu}_n \\ \widehat{v}_n \end{bmatrix} = \widehat{\Omega}^{1/2} \frac{1}{\sqrt{N}} \sum_{n=1}^{[rN]} \mathbf{e}_n \xrightarrow{d^*} B(r). \quad (4.59)$$

Note that the elements of $\widehat{\Omega}$ and $\widehat{\Delta}$ are always multiplicative in the construction. By virtue of (4.59) and the direct application of Lemma 4.2,

$$\begin{aligned} G_{\gamma_0, N}^{-1} \left[\sum_{n=1}^N \dot{f}(\widehat{w}_n, \gamma_0) \widehat{\mu}_n \right] + \widehat{\mathcal{B}}_{vu}^- &\xrightarrow{d^*} \int_0^1 j(r; \gamma_0) dB_u(r) + \begin{bmatrix} \mathbf{0}_{d \times 1} \\ \mathbf{0}_{d \times 1} \\ \Omega_{v_1 u} \mathbf{b}_1 \\ \vdots \\ \Omega_{v_m u} \mathbf{b}_m \end{bmatrix} + \begin{bmatrix} \mathbf{0}_{d \times 1} \\ \mathbf{0}_{d \times 1} \\ \Delta_{v_1 u}^- \mathbf{b}_1 \\ \vdots \\ \Delta_{v_m u}^- \mathbf{b}_m \end{bmatrix} \\ &= \int_0^1 j(r; \gamma_0) dB_u(r) + \begin{bmatrix} \mathbf{0}_{d \times 1} \\ \mathbf{0}_{d \times 1} \\ \Delta_{v_1 u} \mathbf{b}_1 \\ \vdots \\ \Delta_{v_m u} \mathbf{b}_m \end{bmatrix}, \end{aligned} \quad (4.60)$$

where we use $\Omega + \Delta^- = (\Delta + \Delta' - \Sigma) + (\Sigma - \Delta') = \Delta$. Similarly, we have

$$G_{\gamma_0, N}^{-1} \left[\sum_{n=1}^N \dot{f}(\widehat{w}_n, \gamma_0) \dot{f}(\widehat{w}_n, \gamma_0)' \right] G_{\gamma_0, N}^{-1} \xrightarrow{d^*} \int_0^1 j(r; \gamma_0) j(r; \gamma_0)' dr. \quad (4.61)$$

By (4.60) and (4.61), we obtain the theorem. ■

Proof of Theorem 4.4 Without loss of generality, we set $\ell = 1$. Subsequently, note

that

$$\begin{aligned}
 \frac{1}{\sqrt{q_T}} \sum_{t=1}^{[rq_T]} \hat{u}_t^+ &= \frac{1}{\sqrt{q_T}} \sum_{t=1}^{[rq_T]} (u_t - \Omega_{uv} \Omega_{vv}^{-1} v_t) + (\Omega_{uv} \Omega_{vv}^{-1} - \widehat{\Omega}_{uv} \widehat{\Omega}_{vv}^{-1}) \frac{1}{\sqrt{q_T}} \sum_{t=1}^{[rq_T]} v_t \\
 &\quad - \frac{1}{\sqrt{q_T}} \sum_{t=1}^{[rq_T]} (d_t(\widehat{\theta}_T)' \widehat{\tau}_T - d_t(\theta_0)' \tau_0) - \frac{1}{\sqrt{q_T}} \sum_{t=1}^{[rq_T]} s_t' (\widehat{\phi}_T - \phi_0) \\
 &=: VIa + VIb - VIc - VId.
 \end{aligned} \tag{4.62}$$

Assumption 4.2 justifies the use of a functional central limit theorem for linear processes, e.g. Phillips and Solo (1992). Therefore, $VIa \Rightarrow B_{u,v}(r)$ and $\Omega_{u,v}^{-1} \frac{1}{q_T} \sum_{t=1}^{q_T} (\frac{1}{\sqrt{q_T}} \sum_{t=\ell}^t (u_t - \Omega_{uv} \Omega_{vv}^{-1} v_t))^2 \Rightarrow \int_0^1 [W(r)]^2 dr$ by the continuous mapping theorem for functionals. Theorem 4.4 will thus follow if we can show that VIb , VIc and VId are asymptotically negligible.

Because Assumptions 4.1-4.3 are required to hold, Theorem 4.2 implies that $\widehat{\Omega}_T \rightarrow_p \Omega$ (and hence $\|\Omega_{uv} \Omega_{vv}^{-1} - \widehat{\Omega}_{uv} \widehat{\Omega}_{vv}^{-1}\| \rightarrow_p 0$). It follows that $VIb = o_p(1)$. We decompose VIc in three parts:

$$\begin{aligned}
 VIc &= \frac{1}{\sqrt{q_T}} \sum_{t=1}^{[rq_T]} (d_t(\widehat{\theta}_T) - d_t(\theta_0))' \tau_0 + \frac{1}{\sqrt{q_T}} \sum_{t=1}^{[rq_T]} d_t(\theta_0)' (\widehat{\tau}_T - \tau_0) \\
 &\quad + \frac{1}{\sqrt{q_T}} \sum_{t=1}^{[rq_T]} (d_t(\widehat{\theta}_T) - d_t(\theta_0))' (\widehat{\tau}_T - \tau_0) =: VIc^{(1)} + VIc^{(2)} + VIc^{(3)}.
 \end{aligned} \tag{4.63}$$

Note that

$$\begin{aligned}
 |VIc^{(1)}| &= \left| \frac{1}{\sqrt{q_T}} \sum_{t=1}^{[rq_T]} \left(\sum_{i=1}^d (t^{\widehat{\theta}_i} - t^{\theta_{0i}}) \right) \tau_{0i} \right| \\
 &\leq C(\ln q_T) \left(\frac{q_T}{T} \right)^{\theta_L + \frac{1}{2}} \sum_{i=1}^d |\tau_{0i}| T^{\theta_{0i} + \frac{1}{2}} (\widehat{\theta}_i - \theta_{0i}) \left| \frac{1}{q_T} \sum_{t=1}^{q_T} \left(\frac{t}{q_T} \right)^{\theta_{0i}} \right| \\
 &= (\ln q_T) \left(\frac{q_T}{T} \right)^{\theta_L + \frac{1}{2}} O_p(1)
 \end{aligned}$$

by the mean value theorem, Lemma 4.1(i), and $T^{\theta_{0i} + \frac{1}{2}} (\widehat{\theta}_i - \theta_{0i}) = O_p(1)$. By the

Cauchy-Schwartz and triangle inequality, we have

$$\begin{aligned}
|VIC^{(2)}| &= \left| \frac{1}{\sqrt{q_T}} \sum_{t=1}^{\lfloor rq_T \rfloor} [D_{d,q_T}(\theta_0)^{-1} d_t(\theta_0)]' [D_{d,q_T}(\theta_0)(\widehat{\tau}_T - \tau_0)] \right| \\
&= (\ln T) \left(\frac{q_T}{T} \right)^{\theta_L + \frac{1}{2}} \left(\frac{1}{q_T} \sum_{t=1}^{\lfloor rq_T \rfloor} \|D_{d,q_T}(\theta_0)^{-1} d_t(\theta_0)\| \right) \left\| \frac{\sqrt{T}}{\ln T} D_{d,T}(\theta_0)(\widehat{\tau}_T - \tau_0) \right\| \\
&= (\ln T) \left(\frac{q_T}{T} \right)^{\theta_L + \frac{1}{2}} O_p(1),
\end{aligned}$$

where we used $\frac{T^{1/2}}{\ln T} D_{d,T}(\theta_0)(\widehat{\tau}_T - \tau_0) = O_p(1)$ (see Theorem 4.1). Similarly, we bound

$$\begin{aligned}
|VIC^{(3)}| &\leq (\ln T) \left(\frac{q_T}{T} \right)^{\theta_L + \frac{1}{2}} \\
&\quad \times \left(\frac{1}{q_T} \sum_{t=1}^{\lfloor rq_T \rfloor} \|D_{d,q_T}(\theta_0)^{-1} (d_t(\widehat{\theta}_T) - d_t(\theta_0))\| \right) \left\| \frac{\sqrt{T}}{\ln T} D_{d,T}(\theta_0)(\widehat{\tau}_T - \tau_0) \right\|.
\end{aligned}$$

For $i = 1, 2, \dots, d$, we can obtain that $\frac{1}{q_T} \sum_{t=1}^{\lfloor rq_T \rfloor} |q_T^{-\theta_{0i}} (t^{\widehat{\theta}_i} - t^{\theta_{0i}})| \leq C(\ln q_T) \widehat{\theta}_i - \theta_{0i} \left(\frac{1}{q_T} \sum_{t=1}^{\lfloor rq_T \rfloor} \left(\frac{t}{q_T} \right)^{\theta_{0i}} \right) = (\ln q_T) O_p(T^{-(\theta_L + \frac{1}{2})})$ and thus $VIC^{(3)} = o_p(1)$. Overall, $VIC^{(1)}$, $VIC^{(2)}$ and $VIC^{(3)}$ are all three asymptotically negligible under the prerequisite that $(\ln T)(q_T/T)^{\theta_L + \frac{1}{2}} \rightarrow 0$. Finally, term VID . We can write

$$\begin{aligned}
|VID| &= \left| \left(\frac{q_T}{T} \right)^{1/2} \frac{1}{q_T} \sum_{t=1}^{\lfloor rq_T \rfloor} (D_{s,q_T}^{-1} s_t)' D_{s,q_T} D_{s,T}^{-1} [\sqrt{T} D_{s,T}(\widehat{\phi}_T - \phi_0)] \right| \\
&\leq \left(\frac{q_T}{T} \right) \frac{1}{q_T} \sum_{t=1}^{\lfloor rq_T \rfloor} \|D_{s,q_T}^{-1} s_t\| \left\| \sqrt{T} D_{s,T}(\widehat{\phi}_T - \phi_0) \right\| \\
&\leq \left(\frac{q_T}{T} \right) \left(\frac{1}{q_T} \sum_{t=1}^{\lfloor rq_T \rfloor} \|D_{s,q_T}^{-1} s_t\|^2 \right)^{1/2} \left\| \sqrt{T} D_{s,T}(\widehat{\phi}_T - \phi_0) \right\|.
\end{aligned}$$

Using (1) $\frac{1}{q_T} \sum_{t=1}^{\lfloor rq_T \rfloor} \|D_{s,q_T}^{-1} s_t\|^2 = \text{tr} \left(\frac{1}{q_T} \sum_{t=1}^{\lfloor rq_T \rfloor} D_{s,q_T}^{-1} s_t s_t' D_{s,q_T}^{-1} \right) = O_p(1)$, and (2) $\sqrt{T} D_{s,T}(\widehat{\phi}_T - \phi_0) = O_p(1)$, we see that $|VID| = O_p(q_T/T)$. ■

Appendix 4.C Additional Theoretical Results

In this section, we provide some details on Example 4.1 in the main text. Moreover, we comment on the asymptotic invalidity of FMOLS estimator.

Appendix 4.C.1 Limiting Distribution for Example 4.1

Referring to Theorem 4.1, we find

$$\begin{aligned} & \begin{bmatrix} T^{\theta_0 + \frac{1}{2}} & \\ T^{\theta_0 + \frac{1}{2}} \tau_0 \ln(T) & T^{\theta_0 + \frac{1}{2}} \end{bmatrix} \begin{bmatrix} \widehat{\theta}_T - \theta_0 \\ \widehat{\tau}_T - \tau_0 \end{bmatrix} \\ & \Rightarrow \begin{bmatrix} \int_0^1 (\tau_0 r^{\theta_0} \ln(r))^2 dr & \int_0^1 \tau_0 r^{2\theta_0} \ln(r) dr \\ \int_0^1 \tau_0 r^{2\theta_0} \ln(r) dr & \int_0^1 r^{2\theta_0} dr \end{bmatrix}^{-1} \begin{bmatrix} \int_0^1 \tau_0 r^{\theta_0} \ln(r) dB_u \\ \int_0^1 r^{\theta_0} dB_u \end{bmatrix}. \end{aligned}$$

We have to show that the quantity in the RHS is normally distributed with mean and variance as provided in (4.4) of the main paper. Consider an arbitrary vector $\mathbf{c} = [c_1, c_2]'$ and define

$$\begin{aligned} A_{\mathbf{c}} &= \mathbf{c}' \left[\int_0^1 \tau_0 r^{\theta_0} \ln(r) dB_u \int_0^1 r^{\theta_0} dB_u \right] \\ &= \int_0^1 \left[c_1 \tau_0 r^{\theta_0} \ln(r) + c_2 r^{\theta_0} \right] dB_u \stackrel{d}{=} \Omega_{uu}^{1/2} \int_0^1 \left[c_1 \tau_0 r^{\theta_0} \ln(r) + c_2 r^{\theta_0} \right] dW_u. \end{aligned}$$

Gaussianity is preserved under mean square integration (e.g. section 4.6 in Soong (1973)), so $A_{\mathbf{c}}$ it suffices to derive mean and variance. Equation (4.190) in the same reference yields $\mathbb{E}(A_{\mathbf{c}}) = \Omega_{uu}^{1/2} \int_0^1 \left[c_1 \tau_0 r^{\theta_0} \ln(r) + c_2 r^{\theta_0} \right] d\mathbb{B}(W_u) = 0$. Moreover, by (2.16) in Tanaka (2017)

$$\begin{aligned} \mathbb{V}\text{ar}(A_{\mathbf{c}}) &= \Omega_{uu} \int_0^1 \left[c_1 \tau_0 r^{\theta_0} \ln(r) + c_2 r^{\theta_0} \right]^2 dr \\ &= \Omega_{uu} \mathbf{c}' \begin{bmatrix} \int_0^1 (\tau_0 r^{\theta_0} \ln(r))^2 dr & \int_0^1 \tau_0 r^{2\theta_0} \ln(r) dr \\ \int_0^1 \tau_0 r^{2\theta_0} \ln(r) dr & \int_0^1 r^{2\theta_0} dr \end{bmatrix} \mathbf{c}. \end{aligned}$$

c was arbitrary and therefore

$$\begin{bmatrix} \int_0^1 \tau_0 r^{\theta_0} \ln(r) dB_u \\ \int_0^1 r^{\theta_0} dB_u \end{bmatrix} \sim N\left(\mathbf{0}, \Omega_{uu} \begin{bmatrix} \int_0^1 (\tau_0 r^{\theta_0} \ln(r))^2 dr & \int_0^1 \tau_0 r^{2\theta_0} \ln(r) dr \\ \int_0^1 \tau_0 r^{2\theta_0} \ln(r) dr & \int_0^1 r^{2\theta_0} dr \end{bmatrix}\right).$$

Finally, use $\int_0^1 (r^{\theta_0} \ln(r))^2 dr = \frac{2}{(2\theta_0+1)^3}$, $\int_0^1 r^{2\theta_0} \ln(r) dr = -\frac{1}{(2\theta_0+1)^2}$, and basic linear algebra to recover the claim of Example 4.1.

Appendix 4.C.2 FMOLS Estimator

We here comment on the asymptotic properties of the FMOLS estimator. To be specific, we analyse the asymptotic properties of $\tilde{D}_{\theta_0, T} \begin{bmatrix} \hat{\tau}_T^+ - \tau_0 \\ \hat{\phi}_T^+ - \phi_0 \end{bmatrix}$, with $\tilde{D}_{\theta_0, T} = \sqrt{T} \begin{bmatrix} D_{d,T}(\theta_0) & \mathbf{0}_{d \times p} \\ \mathbf{0}_{p \times d} & D_{s,T} \end{bmatrix}$ and

$$\begin{bmatrix} \hat{\tau}_T^+ \\ \hat{\phi}_T^+ \end{bmatrix} = \left(\sum_{t=1}^T z_t(\hat{\theta}_T) z_t(\hat{\theta}_T)' \right)^{-1} \left(\sum_{t=1}^T z_t(\hat{\theta}_T) y_t^+ - A^* \right),$$

where y_t^+ and A^* are the usual second-order bias corrections. That is, $y_t^+ = y_t - \hat{\Omega}_{uv} \hat{\Omega}_{vv}^{-1} \Delta x_t$ and $A^* = [\mathbf{0}'_{d \times 1}, A_1^*, \dots, A_m^*]'$ with $A_i^* = \hat{\Delta}_{viu}^+ \left[T, 2 \sum_{t=1}^T x_{it}, \dots, p_i \sum_{t=1}^T x_{it}^{p_i-1} \right]'$ and $\hat{\Delta}_{viu}^+$ is the i^{th} row of $\hat{\Delta}_{vu}^+ = \hat{\Delta}_{vu} - \hat{\Delta}_{vv} \hat{\Omega}_{vv}^{-1} \hat{\Omega}_{vu}$ ($i = 1, 2, \dots, m$). If the converge speed of $\hat{\theta}_T$ is sufficiently fast, then its estimation error is asymptotically negligible and the limiting distribution of $\tilde{D}_{\theta_0, T} \begin{bmatrix} \hat{\tau}_T^+ - \tau_0 \\ \hat{\phi}_T^+ - \phi_0 \end{bmatrix}$ is mixed normal.

We now focus on the limiting distribution. By linear algebra manipulations, we find

$$\tilde{D}_{\theta_0, T} \begin{bmatrix} \hat{\tau}_T^+ - \tau_0 \\ \hat{\phi}_T^+ - \phi_0 \end{bmatrix} = \left(\tilde{D}_{\theta_0, T}^{-1} \sum_{t=1}^T z_t(\hat{\theta}_T) z_t(\hat{\theta}_T)' \tilde{D}_{\theta_0, T}^{-1} \right)^{-1} \tilde{D}_{\theta_0, T}^{-1} \left[\sum_{t=1}^T z_t(\hat{\theta}_T) \tilde{u}_t^+ - A^* \right], \quad (4.64)$$

where $\tilde{u}_t^+ = (z_t(\theta_0) - z_t(\hat{\theta}_T))' \begin{bmatrix} \tau_0 \\ \phi_0 \end{bmatrix} + u_t - \hat{\Omega}_{uv} \hat{\Omega}_{vv}^{-1} \Delta x_t$. We will discuss $\tilde{D}_{\theta_0, T}^{-1} \sum_{t=1}^T z_t(\hat{\theta}_T) z_t(\hat{\theta}_T)' \tilde{D}_{\theta_0, T}^{-1}$ and $\tilde{D}_{\theta_0, T}^{-1} \left[\sum_{t=1}^T z_t(\hat{\theta}_T) \tilde{u}_t^+ - A^* \right]$ separately after having enumerate several intermediate results.

Lemma 4.5

Define $\tilde{j}(r; \theta_0) = [d(r; \theta_0)', B'_{(1)}(r), \dots, B'_{(m)}(r)]'$ and $B_{u,v} = B_u - \Omega_{uv} \Omega_{vv}^{-1} B_v$. Then, under Assumptions 4.1-4.3, we have

- (i) $\tilde{D}_{\theta_0, T}^{-1} \sum_{t=1}^T z_t(\hat{\theta}_T) z_t(\hat{\theta}_T)' \tilde{D}_{\theta_0, T}^{-1} \Rightarrow \int_0^1 \tilde{j}(r; \theta_0) \tilde{j}(r; \theta_0)' dr,$
- (ii) $\tilde{D}_{\theta_0, T}^{-1} \left[\sum_{t=1}^T z_t(\theta_0) (u_t - \hat{\Omega}_{uv} \hat{\Omega}_{vv}^{-1} v_t) - A^* \right] \Rightarrow \int_0^1 \tilde{j}(r; \theta_0) dB_{u,v}(r),$
- (iii) $\tilde{D}_{\theta_0, T}^{-1} \sum_{t=1}^T z_t(\theta_0) (z_t(\hat{\theta}_T) - z_t(\theta_0))' \begin{bmatrix} \tau_0 \\ \phi_0 \end{bmatrix} = O_p(\ln T),$
- (iv) $\sum_{t=1}^{b_T} \tilde{D}_{\theta_0, b_T}^{-1} (z_t(\hat{\theta}_T) - z_t(\theta_0)) (z_t(\hat{\theta}_T) - z_t(\theta_0))' \begin{bmatrix} \tau_0 \\ \phi_0 \end{bmatrix} = O_p((\ln T)^2 T^{-(\theta_L + \frac{1}{2})}),$
- (v) $\tilde{D}_{\theta_0, T}^{-1} \sum_{t=1}^T (z_t(\hat{\theta}_T) - z_t(\theta_0)) (u_t - \hat{\Omega}_{uv} \hat{\Omega}_{vv}^{-1} v_t) = o_p(1).$

Proof (i) We can always add and subtract such that the LHS of (i) reads

$$\begin{aligned} \tilde{D}_{\theta_0, T}^{-1} \sum_{t=1}^T z_t(\hat{\theta}_T) z_t(\hat{\theta}_T)' \tilde{D}_{\theta_0, T}^{-1} &= \tilde{D}_{\theta_0, T}^{-1} \sum_{t=1}^T z_t(\theta_0) z_t(\theta_0)' \tilde{D}_{\theta_0, T}^{-1} \\ &+ \left(\tilde{D}_{\theta_0, T}^{-1} \sum_{t=1}^T z_t(\hat{\theta}_T) z_t(\hat{\theta}_T)' \tilde{D}_{\theta_0, T}^{-1} - \tilde{D}_{\theta_0, T}^{-1} \sum_{t=1}^T z_t(\theta_0) z_t(\theta_0)' \tilde{D}_{\theta_0, T}^{-1} \right). \end{aligned} \quad (4.65)$$

Lemma 4.2(iii) implies that the first term in the RHS of (4.65) converges to $\int_0^1 \tilde{j}(r; \theta_0) \tilde{j}(r; \theta_0)' dr$. It remains to show that the term in brackets vanishes. By the Cauchy-Schwarz inequality, we have

$$\begin{aligned} &\left\| \tilde{D}_{\theta_0, T}^{-1} \sum_{t=1}^T z_t(\hat{\theta}_T) z_t(\hat{\theta}_T)' \tilde{D}_{\theta_0, T}^{-1} - \tilde{D}_{\theta_0, T}^{-1} \sum_{t=1}^T z_t(\theta_0) z_t(\theta_0)' \tilde{D}_{\theta_0, T}^{-1} \right\| \\ &\leq \sum_{t=1}^T \left\| \tilde{D}_{\theta_0, T}^{-1} (z_t(\hat{\theta}_T) - z_t(\theta_0)) \right\|^2 + 2 \sum_{t=1}^T \left\| \tilde{D}_{\theta_0, T}^{-1} z_t(\theta_0) \right\| \left\| \tilde{D}_{\theta_0, T}^{-1} (z_t(\hat{\theta}_T) - z_t(\theta_0)) \right\| \\ &\leq \sum_{t=1}^T \left\| \tilde{D}_{\theta_0, T}^{-1} (z_t(\hat{\theta}_T) - z_t(\theta_0)) \right\|^2 + 2 \left[\sum_{t=1}^T \left\| \tilde{D}_{\theta_0, T}^{-1} z_t(\theta_0) \right\|^2 \right]^{1/2} \left[\sum_{t=1}^T \left\| \tilde{D}_{\theta_0, T}^{-1} (z_t(\hat{\theta}_T) - z_t(\theta_0)) \right\|^2 \right]^{1/2}. \end{aligned}$$

We have $\sum_{t=1}^T \left\| \tilde{D}_{\theta_0, T}^{-1} z_t(\theta_0) \right\|^2 = \text{tr} \left(\sum_{t=1}^T \tilde{D}_{\theta_0, T}^{-1} z_t(\theta_0) z_t(\theta_0)' \tilde{D}_{\theta_0, T}^{-1} \right) \Rightarrow \text{tr} \left(\int_0^1 \tilde{j}(r; \theta_0) \tilde{j}(r; \theta_0)' dr \right)$. Next note that $\sum_{t=1}^T \left\| \tilde{D}_{\theta_0, T}^{-1} (z_t(\hat{\theta}_T) - z_t(\theta_0)) \right\|^2 =$

$\frac{1}{T} \sum_{t=1}^T \|\mathbf{D}_{d,T}(\boldsymbol{\theta}_0)^{-1}(\mathbf{d}_t(\widehat{\boldsymbol{\theta}}_T) - \mathbf{d}_t(\boldsymbol{\theta}_0))\|^2$. A typical contribution to the latter sum of norms can be bound by

$$\begin{aligned} \frac{1}{T} \sum_{t=1}^T [T^{-\theta_{0i}}(t^{\widehat{\theta}_i} - t^{\theta_{0i}})]^2 &\leq C(\widehat{\theta}_i - \theta_{0i}) \frac{1}{T} \sum_{t=1}^T \left(\frac{t}{T}\right)^{2\theta_{0i}} (\ln t)^2 \\ &\leq CT^{-2(\theta_{0i}-\frac{1}{2})}(\ln T)^2 \left[T^{\theta_{0i}+\frac{1}{2}}(\widehat{\theta}_i - \theta_{0i})\right]^2 \sup_{\theta_L \leq \theta \leq \theta_U} \left|\frac{1}{T} \sum_{t=1}^T \left(\frac{t}{T}\right)^{2\theta}\right| = o_p(1), \end{aligned} \quad (4.66)$$

where we used the mean-value theorem and Lemma 4.1(i). The claim follows.

(ii) $\widehat{\boldsymbol{\Omega}}_{uv}$ and $\widehat{\boldsymbol{\Omega}}_{vv}$ are consistently estimating $\boldsymbol{\Omega}_{uv}$ and $\boldsymbol{\Omega}_{vv}$, respectively (Theorem 4.2). It therefore suffices to look at $\widetilde{\mathbf{D}}_{\boldsymbol{\theta}_0,T}^{-1} \sum_{t=1}^T \mathbf{z}_t(\boldsymbol{\theta}_0)(u_t - \boldsymbol{\Omega}_{uv}\boldsymbol{\Omega}_{vv}^{-1}\mathbf{v}_t)$ and $\widetilde{\mathbf{D}}_{\boldsymbol{\theta}_0,T}^{-1}\mathbf{A}^*$. Lemma 4.2(ii) with $u_t^+ = u_t - \boldsymbol{\Omega}_{uv}\boldsymbol{\Omega}_{vv}^{-1}\mathbf{v}_t$ instead of u_t gives the limiting result $\frac{1}{\sqrt{T}} \sum_{t=1}^T (x_{it}/\sqrt{T})^j u_t^+ \Rightarrow \int_0^1 B_{v_i}^j(r)dB_{u,v}(r) + j\boldsymbol{\Delta}_{v_iu}^+ \int_0^1 B_{v_i}^{j-1}(r)dr$, which implies

$$\widetilde{\mathbf{D}}_{\boldsymbol{\theta}_0,T}^{-1} \sum_{t=1}^T \mathbf{z}_t(\boldsymbol{\theta}_0)(u_t - \boldsymbol{\Omega}_{uv}\boldsymbol{\Omega}_{vv}^{-1}\mathbf{v}_t) \Rightarrow \int_0^1 \widetilde{\mathbf{j}}(r; \gamma_0)dB_{u,v}(r) + \widetilde{\mathcal{B}}_{vu}^+, \quad (4.67)$$

where $\widetilde{\mathcal{B}}_{vu}^+ = [\mathbf{0}'_{d \times 1}, \mathbf{b}'_1 \boldsymbol{\Delta}_{v_1u}^+, \dots, \mathbf{b}'_m \boldsymbol{\Delta}_{v_mu}^+]'$. The term $-\widetilde{\mathbf{D}}_{\boldsymbol{\theta}_0,T}^{-1}\mathbf{A}^*$ is constructed to asymptotically cancel out the term $\widetilde{\mathcal{B}}_{vu}^+$ in the RHS of (4.67). (iii) Using $\mathbf{z}_t(\widehat{\boldsymbol{\theta}}_T) - \mathbf{z}_t(\boldsymbol{\theta}_0) = \begin{bmatrix} \mathbf{d}_t(\widehat{\boldsymbol{\theta}}_T) - \mathbf{d}_t(\boldsymbol{\theta}_0) \\ 0 \end{bmatrix}$, we have

$$\widetilde{\mathbf{D}}_{\boldsymbol{\theta}_0,T}^{-1} \sum_{t=1}^T \mathbf{z}_t(\boldsymbol{\theta}_0)(\mathbf{z}_t(\widehat{\boldsymbol{\theta}}_T) - \mathbf{z}_t(\boldsymbol{\theta}_0))' \begin{bmatrix} \boldsymbol{\tau}_0 \\ \boldsymbol{\phi}_0 \end{bmatrix} = \begin{bmatrix} T^{-1/2} \sum_{t=1}^T \mathbf{D}_{d,T}(\boldsymbol{\theta}_0)^{-1} \mathbf{d}_t(\boldsymbol{\theta}_0)(\mathbf{d}_t(\widehat{\boldsymbol{\theta}}_T) - \mathbf{d}_t(\boldsymbol{\theta}_0))' \boldsymbol{\tau}_0 \\ T^{-1/2} \sum_{t=1}^T \mathbf{D}_{s,T}^{-1} \mathbf{s}_t(\mathbf{d}_t(\widehat{\boldsymbol{\theta}}_T) - \mathbf{d}_t(\boldsymbol{\theta}_0))' \boldsymbol{\tau}_0 \end{bmatrix}.$$

The typical elements in the vector on the RHS are of the form $\frac{1}{\sqrt{T}} \sum_{t=1}^T \left(\frac{t}{T}\right)^{\theta_{0i}} \sum_{k=1}^d \tau_{0k}(t^{\widehat{\theta}_k} - t^{\theta_{0k}})$ or $\frac{1}{\sqrt{T}} \sum_{t=1}^T \left(\frac{x_{it}}{\sqrt{T}}\right)^j \sum_{k=1}^d \tau_{0k}(t^{\widehat{\theta}_k} - t^{\theta_{0k}})$. We show that both contributions are $O_p(\ln T)$. By the mean-value theorem and Lemma 4.1(i),

$$\begin{aligned} \left| \frac{1}{\sqrt{T}} \sum_{t=1}^T \left(\frac{t}{T}\right)^{\theta_{0i}} \sum_{k=1}^d \tau_{0k}(t^{\widehat{\theta}_k} - t^{\theta_{0k}}) \right| &\leq \left| \frac{1}{\sqrt{T}} \sum_{k=1}^d \tau_{0k} \sum_{t=1}^T \left(\frac{t}{T}\right)^{\theta_{0i}} t^{\theta_{0k}} (t^{\widehat{\theta}_k - \theta_{0k}} - 1) \right| \\ &\leq C \sum_{k=1}^d |\tau_{0k}| T^{\theta_{0k}+\frac{1}{2}} (\widehat{\theta}_k - \theta_{0k}) \left| \frac{1}{T} \sum_{t=1}^T \left(\frac{t}{T}\right)^{\theta_{0i}+\theta_{0k}} \ln t \right| \\ &\leq C(\ln T) \sum_{k=1}^d |\tau_{0k}| T^{\theta_{0k}+\frac{1}{2}} (\widehat{\theta}_k - \theta_{0k}) \left| \left[\frac{1}{T} \sum_{t=1}^T \left(\frac{t}{T}\right)^{\theta_{0i}+\theta_{0k}} \right] \right| \\ &= O_p(\ln T). \end{aligned} \quad (4.68)$$

Similarly, from the mean-value theorem and Cauchy-Schwartz inequality, we see that

$$\begin{aligned}
 & \left| \frac{1}{\sqrt{T}} \sum_{t=1}^T \left(\frac{x_{it}}{\sqrt{T}} \right)^j \sum_{k=1}^d \tau_{0k} (t^{\widehat{\theta}_k} - t^{\theta_{0k}}) \right| \leq \left| \frac{1}{\sqrt{T}} \sum_{k=1}^d \tau_{0k} \sum_{t=1}^T \left(\frac{x_{it}}{\sqrt{T}} \right)^j t^{\theta_{0k}} (t^{\widehat{\theta}_k - \theta_{0k}} - 1) \right| \\
 & \leq C \sum_{k=1}^d |\tau_{0k}| |T^{\theta_{0k} + \frac{1}{2}} (\widehat{\theta}_k - \theta_{0k})| \frac{1}{T} \sum_{t=1}^T \left| \frac{x_{it}}{\sqrt{T}} \right|^j \left(\frac{t}{T} \right)^{\theta_{0k}} \ln t \\
 & \leq C(\ln T) \sum_{k=1}^d |\tau_{0k}| |T^{\theta_{0k} + \frac{1}{2}} (\widehat{\theta}_k - \theta_{0k})| \left[\frac{1}{T} \sum_{t=1}^T \left(\frac{x_{it}}{\sqrt{T}} \right)^{2j} \right]^{1/2} \left[\frac{1}{T} \sum_{t=1}^T \left(\frac{t}{T} \right)^{2\theta_L} \right]^{1/2}. \quad (4.69)
 \end{aligned}$$

From (4.68) and (4.69) we conclude that $\widetilde{D}_{\theta_0, T}^{-1} \sum_{t=1}^T z_t(\theta_0) (z_t(\widehat{\theta}_T) - z_t(\theta_0))' \begin{bmatrix} \tau_0 \\ \phi_0 \end{bmatrix} = O_p(\ln T)$. (iv) Use $z_t(\widehat{\theta}_T) - z_t(\theta_0) = \begin{bmatrix} d_t(\widehat{\theta}_T) - d_t(\theta_0) \\ 0 \end{bmatrix}$ to obtain

$$\begin{aligned}
 & \widetilde{D}_{\theta_0, T}^{-1} \sum_{t=1}^T (z_t(\widehat{\theta}_T) - z_t(\theta_0)) (z_t(\widehat{\theta}_T) - z_t(\theta_0))' \begin{bmatrix} \tau_0 \\ \phi_0 \end{bmatrix} \\
 & = \begin{bmatrix} D_{d, T}(\theta_0)^{-1} \frac{1}{\sqrt{T}} \sum_{t=1}^T (d_t(\widehat{\theta}_T) - d_t(\theta_0)) (d_t(\widehat{\theta}_T) - d_t(\theta_0))' \tau_0 \\ \mathbf{0}_{p \times 1} \end{bmatrix}.
 \end{aligned}$$

For any $i \in \{1, 2, \dots, d\}$, the norm of the i^{th} component of the nonzero vector is

$$\begin{aligned}
 & \left| \sum_{k=1}^d \tau_{0k} \frac{1}{T^{\theta_{0i} + 1/2}} \sum_{t=1}^T (t^{\widehat{\theta}_i} - t^{\theta_{0i}}) (t^{\widehat{\theta}_k} - t^{\theta_{0k}}) \right| \\
 & \leq \sum_{k=1}^d |\tau_{0k}| \frac{1}{T^{\theta_{0i} + 1/2}} \sum_{t=1}^T t^{\theta_{0i} + \theta_{0k}} |t^{\widehat{\theta}_i - \theta_{0i}} - 1| |t^{\widehat{\theta}_k - \theta_{0k}} - 1| \\
 & \leq C \sum_{k=1}^d |\tau_{0k}| |\widehat{\theta}_i - \theta_{0i}| |\widehat{\theta}_k - \theta_{0k}| \frac{1}{T^{\theta_{0i} + 1/2}} \sum_{t=1}^T t^{\theta_{0i} + \theta_{0k}} (\ln t)^2 \\
 & \leq C(\ln T)^2 T^{-(\theta_L + \frac{1}{2})} |T^{\theta_{0i} + \frac{1}{2}} (\widehat{\theta}_i - \theta_{0i})| \sum_{k=1}^d |\tau_{0k}| |T^{\theta_{0k} + \frac{1}{2}} (\widehat{\theta}_k - \theta_{0k})| \left[\frac{1}{T} \sum_{t=1}^T \left(\frac{t}{T} \right)^{\theta_{0i} + \theta_{0k}} \right] \\
 & = O_p \left(\frac{(\ln T)^2}{T^{\theta_L + \frac{1}{2}}} \right).
 \end{aligned}$$

(v) By similar steps as before, and invoking Theorem 4.2, it is easy to show that it suffices to bound $T^{-(\theta_{0i} + \frac{1}{2})} \sum_{t=1}^T (t^{\widehat{\theta}_i} - t^{\theta_{0i}}) (u_t - \Omega_{uv} \Omega_{vv}^{-1} v_t)$. Writing $u_t^+ = u_t - \Omega_{uv} \Omega_{vv}^{-1} v_t$

(as before), we have

$$\begin{aligned}
T^{-(\theta_{0i} + \frac{1}{2})} \sum_{t=1}^T (\widehat{t}^{\widehat{\theta}_i} - t^{\theta_{0i}}) u_t^+ &= \frac{1}{\sqrt{T}} \sum_{t=1}^T \left(\frac{t}{T} \right)^{\theta_{0i}} (t^{\widehat{\theta}_i - \theta_{0i}} - 1) u_t^+ \\
&= (\widehat{\theta}_i - \theta_{0i}) \frac{1}{\sqrt{T}} \sum_{t=1}^T (\ln t) \left(\frac{t}{T} \right)^{\theta_{0i}} u_t^+ + o_p(1) \\
&= T^{-(\theta_{0i} + \frac{1}{2})} \left[T^{\theta_{0i} + \frac{1}{2}} (\widehat{\theta}_i - \theta_{0i}) \right] \frac{1}{\sqrt{T}} \sum_{t=1}^T \left(\ln \frac{t}{T} \right) \left(\frac{t}{T} \right)^{\theta_{0i}} u_t^+ \\
&\quad + T^{-(\theta_{0i} + \frac{1}{2})} \left[T^{\theta_{0i} + \frac{1}{2}} (\widehat{\theta}_i - \theta_{0i}) \right] (\ln T) \frac{1}{\sqrt{T}} \sum_{t=1}^T \left(\frac{t}{T} \right)^{\theta_{0i}} u_t^+ + o_p(1) \\
&= \frac{1}{T^{\theta_{0i} + \frac{1}{2}}} O_p(1) + \frac{\ln T}{T^{\theta_{0i} + \frac{1}{2}}} O_p(1).
\end{aligned}$$

This establishes (v). ■

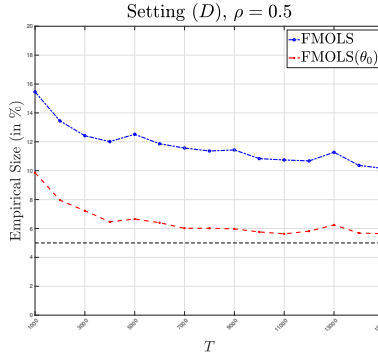


Figure 4.3: Empirical size of feasible and infeasible FMOLS estimators, see the note for Table 4.2.

The current upper bounds in the lemma above suggest that the RHS of (4.64) does not converge to a Gaussian mixture limiting distribution when θ_0 is unknown and has to be estimated. An additional simulation study was conducted to verify this claim. That is, we extend the simulation study on the Monte Carlo results for testing $H_0 : \phi_2 = 0$ versus $H_a : \phi_2 \neq 0$ to higher sample sizes. We consider serial correlation setting (D) and $\rho = 0.50$ as in the last column of Table 4.2. For sample sizes as large as 15000, the empirical size

of feasible FMOLS estimator that relies on $\widehat{\theta}_T$ fluctuates around 11% (Figure 4.3). This indeed points towards a lack of asymptotic validity. On the contrary, FMOLS(θ_0) yields an empirical size close to 5%.

Appendix 4.D Simulation Results on Confidence Intervals for θ

Table 4.5 reports the empirical coverage and average confidence interval (CI) length of 95% confidence intervals for θ . The coverage is always below the desired nominal level of 95%. Coverage can drop as low as 54% when the sample size is small ($T = 100$) and the high serial correlation scenario (D) is used. This lack of coverage is caused by the imprecise estimation of the long-run variance (LRV) matrices. If we provide the true values of the LRV matrices, see the rows labeled Coverage(Ω), then coverage is almost exactly 95% throughout all designs. As expected, the average width of the CIs decreases with sample size.

Table 4.5: Simulation results on the confidence intervals for θ . We report the empirical coverage, the coverage(Ω) computed with the true LRVs, and average length of 95% confidence intervals. All computations use simulated inference, see Section 4.3.2.

	(A)			(B)			(C)			(D)		
ρ	0.00	0.25	0.50	0.00	0.25	0.50	0.00	0.25	0.50	0.00	0.25	0.50
$T = 100$												
Coverage	93.08	92.65	92.20	84.64	84.77	84.24	76.90	76.85	76.66	54.54	54.74	55.35
Coverage(Ω)	94.55	94.64	95.44	94.49	94.38	94.07	94.72	94.59	94.45	95.66	95.76	96.14
Length	0.66	0.70	0.81	0.87	0.89	0.92	1.14	1.14	1.16	1.76	1.79	1.76
$T = 200$												
Coverage	94.22	93.85	93.36	89.41	89.00	88.71	84.88	84.37	84.79	68.76	69.04	68.33
Coverage(Ω)	94.98	94.89	94.99	95.22	94.80	94.34	95.06	95.04	95.07	95.61	95.53	95.97
Length	0.12	0.13	0.15	0.17	0.17	0.18	0.24	0.24	0.24	0.39	0.39	0.39
$T = 500$												
Coverage	94.55	94.44	94.35	91.68	91.61	91.60	90.22	90.02	89.42	82.49	82.33	81.90
Coverage(Ω)	94.94	95.03	94.89	94.75	94.64	94.40	95.01	95.02	94.51	95.07	95.26	95.51
Length	0.01	0.01	0.02	0.02	0.02	0.02	0.03	0.03	0.03	0.06	0.06	0.06

Note: The DGP is $y_t = \tau_1 + \tau_2 t + \tau_3 t^2 + \phi_1 x_t + \phi_2 x_t^2 + u_t$ (numerical values are given on page 170) with $x_t = \sum_{s=1}^t v_s$. The innovations follow $\begin{bmatrix} u_t \\ v_t \end{bmatrix} = A \begin{bmatrix} u_{t-1} \\ v_{t-1} \end{bmatrix} + \begin{bmatrix} \eta_t \\ \epsilon_t \end{bmatrix}$ with $\begin{bmatrix} \eta_t \\ \epsilon_t \end{bmatrix} \stackrel{i.i.d.}{\sim} N\left(0, \begin{bmatrix} 1 & \rho \\ \rho & 1 \end{bmatrix}\right)$. $A = HLH'$ with $H = U(U'U)^{-1/2}$ and U being a (2×2) matrix of uniformly distributed random variables on $[0, 1]$. We consider four specifications for L : (A) $L = \text{diag}[0, 0]$, (B) $L = \text{diag}[0.5, 0.3]$, (C) $L = \text{diag}[0.7, 0.5]$, and (D) $L = \text{diag}[0.9, 0.7]$.

Appendix 4.E Further Empirical Results

Appendix 4.E.3 Unit Root Tests

Table 4.6: The t-statistics for the ADF and DF-GLS unit root tests. The columns with header ‘const’ and ‘const & trend’ refer to the inclusion of only an intercept or both intercept and linear trend. Rejection of the unit root hypothesis at a 10% and 5% level are indicated with one and two stars, respectively.

	ADF				DF-GLS			
	const		const & trend		const		const & trend	
	GDP	CO ₂	GDP	CO ₂	GDP	CO ₂	GDP	CO ₂
Australia	0.287	-2.549	-2.050	-1.986	2.046	1.379	-1.577	-0.732
Austria	-0.055	-2.118	-1.943	-2.738	1.478	-1.143	-1.655	-2.718*
Belgium	0.153	-2.336	-1.705	-2.818	2.041	-0.794	-1.287	-2.644
Canada	-0.575	-1.133	-2.020	-1.120	1.117	0.874	-1.894	-0.387
Denmark	-0.235	-2.446	-2.326	-0.136	1.393	0.410	-1.505	0.084
Finland	-0.362	-1.327	-2.315	-3.248*	0.420	-0.076	-1.155	-3.217**
France	-0.557	-2.438	-1.823	-1.858	1.087	-0.267	-1.470	-1.212
Germany	-0.374	-3.099**	-2.767	-3.971**	1.195	-0.726	-2.474	-2.080
Italy	-0.252	-1.546	-1.759	-1.987	1.213	0.354	-1.240	-1.860
Japan	0.010	-0.862	-1.733	-0.941	1.382	0.504	-1.272	-0.878
Netherlands	-0.106	-1.629	-2.247	-3.106	1.378	0.213	-1.679	-2.818*
Norway	-0.680	-2.044	-2.064	-2.318	0.749	0.331	-1.017	-1.292
Portugal	-1.432	-0.455	-1.697	-1.676	-0.708	0.593	-0.741	-1.923
Spain	0.402	-1.243	-1.354	-1.994	1.487	0.959	-1.077	-2.014
Sweden	-0.789	-2.075	-2.289	-1.625	0.258	0.180	-1.513	-0.968
Switzerland	-1.093	-1.963	-2.785	-1.989	2.272	0.368	-2.447	-1.237
UK	-0.179	-0.721	-1.262	-0.402	2.446	-0.622	-0.608	-0.013
USA	-0.349	-2.055	-2.871	-1.322	2.409	-0.101	-2.708*	-0.812

Note: Asterisks denote rejection of the null hypothesis at the ***1%, **5%, and *10% significance level.

Appendix 4.E.4 Perron and Yabu (2009) Test for Deterministic Trend Coefficient

The Perron and Yabu (2009) test is used to test for the presence of a deterministic trend function in the log per capita GDP series, see Table 4.7. The test allows for integrated or stationary errors. The details of the procedure can be

found on page 61 of Perron and Yabu (2009). The asymptotic distribution of this test statistic is standard normal (quantiles are $z_{0.95} = 1.645$, $z_{0.975} = 1.96$, and $z_{0.995} = 2.58$).

Table 4.7: Perron and Yabu (2009) test statistic for each of the 18 countries.

	\widehat{PY}		\widehat{PY}		\widehat{PY}
Australia	3.17	France	2.41	Portugal	2.16
Austria	2.19	Germany	1.91	Spain	2.31
Belgium	3.52	Italy	2.11	Sweden	7.12
Canada	3.33	Japan	2.93	Switzerland	3.91
Denmark	5.58	Netherlands	2.27	UK	3.60
Finland	4.27	Norway	5.85	USA	4.12

Appendix 4.E.5 Other Figures

As a robustness check, in addition to Belgium (see Figure 4.2), we also provide overviews for Austria and Finland in Figures 4.4 and 4.5. Moreover, the fit for the recommended model (M4) is displayed in Figure 4.6. Finally, we show the residual series for Models (M1)-(M4) in Figures 4.7 to 4.10.

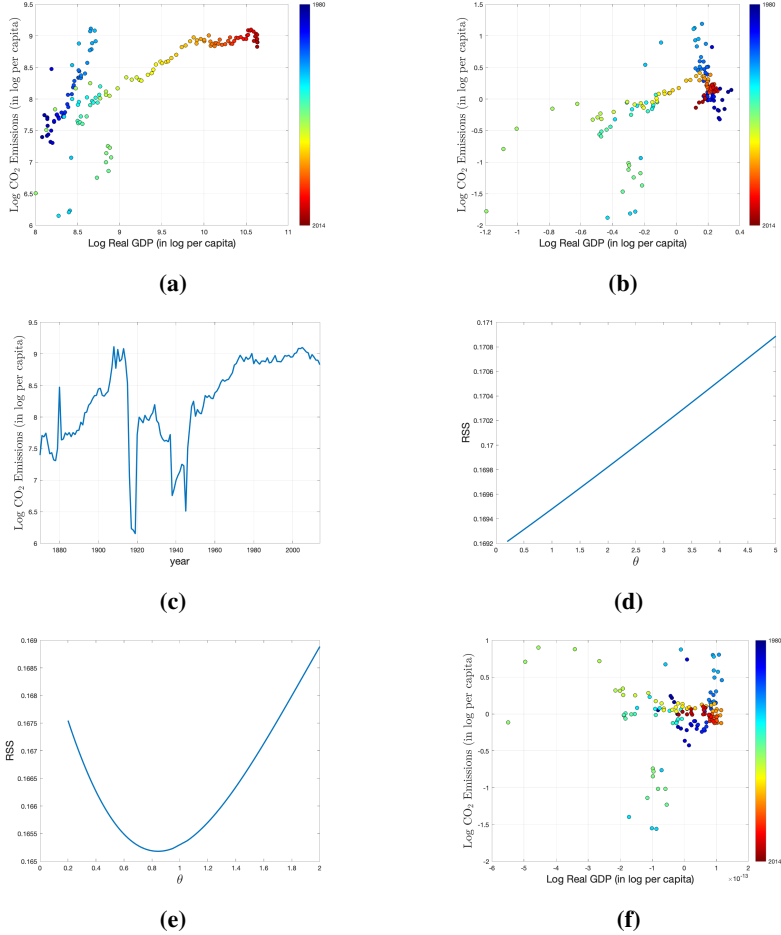


Figure 4.4: Overview graphs for Austria over 1870-2014. **(a)** log(GDP) versus log(CO₂) (both per capita). **(b)** As subfigure (a) but using detrended variables. **(c)** The log per capita CO₂ emissions time series for Austria. **(d)** The residual sum of squares (RSS) for the nonlinear model specification $y_t = \tau_1 + \tau_2 t + \phi_1 x_t + \phi_2 x_t^\theta + u_t$ for various values of θ . **(e)** The RSS as a function of θ for the flexible nonlinear trend specification $y_t = \tau_1 + \tau_2 t + \tau_3 t^\theta + \phi x_t + u_t$. **(f)** The relation between x_t and y_t after partialling out the constant, linear trend, and flexible deterministic trend.

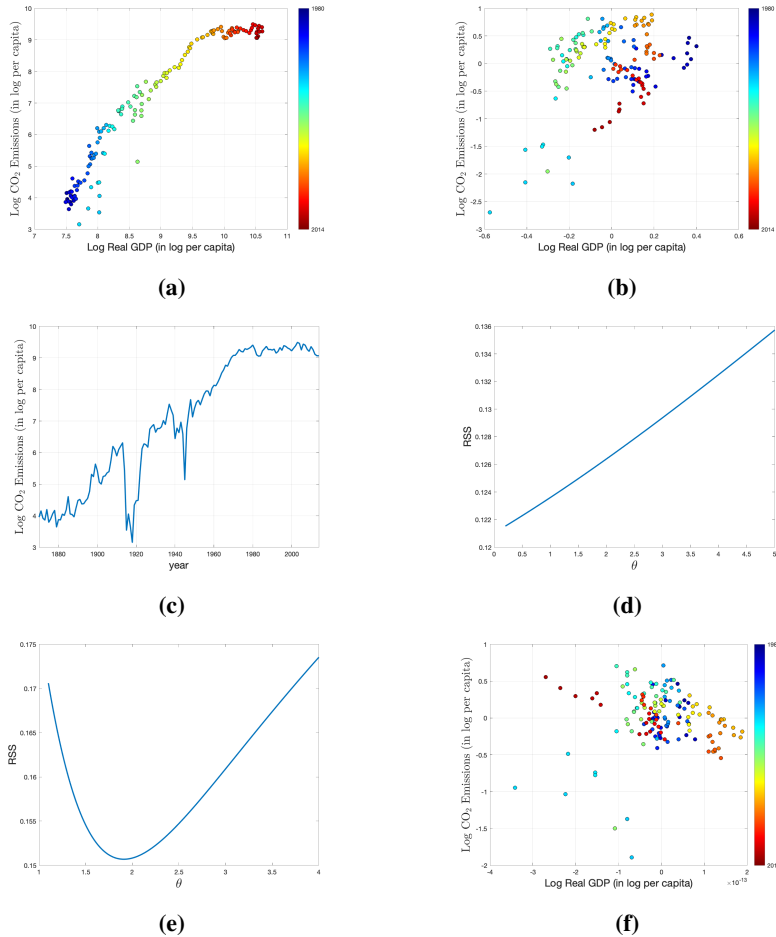


Figure 4.5: Overview graphs for Finland over 1870-2014. **(a)** $\log(\text{GDP})$ versus $\log(\text{CO}_2)$ (both per capita). **(b)** As subfigure (a) but using detrended variables. **(c)** The log per capita CO₂ emissions time series for Finland. **(d)** The residual sum of squares (RSS) for the nonlinear model specification $y_t = \tau_1 + \tau_2 t + \phi_1 x_t + \phi_2 x_t^\theta + u_t$ for various values of θ . **(e)** The RSS as a function of θ for the flexible nonlinear trend specification $y_t = \tau_1 + \tau_2 t + \tau_3 t^\theta + \phi x_t + u_t$. **(f)** The relation between x_t and y_t after partialling out the constant, linear trend, and flexible deterministic trend.

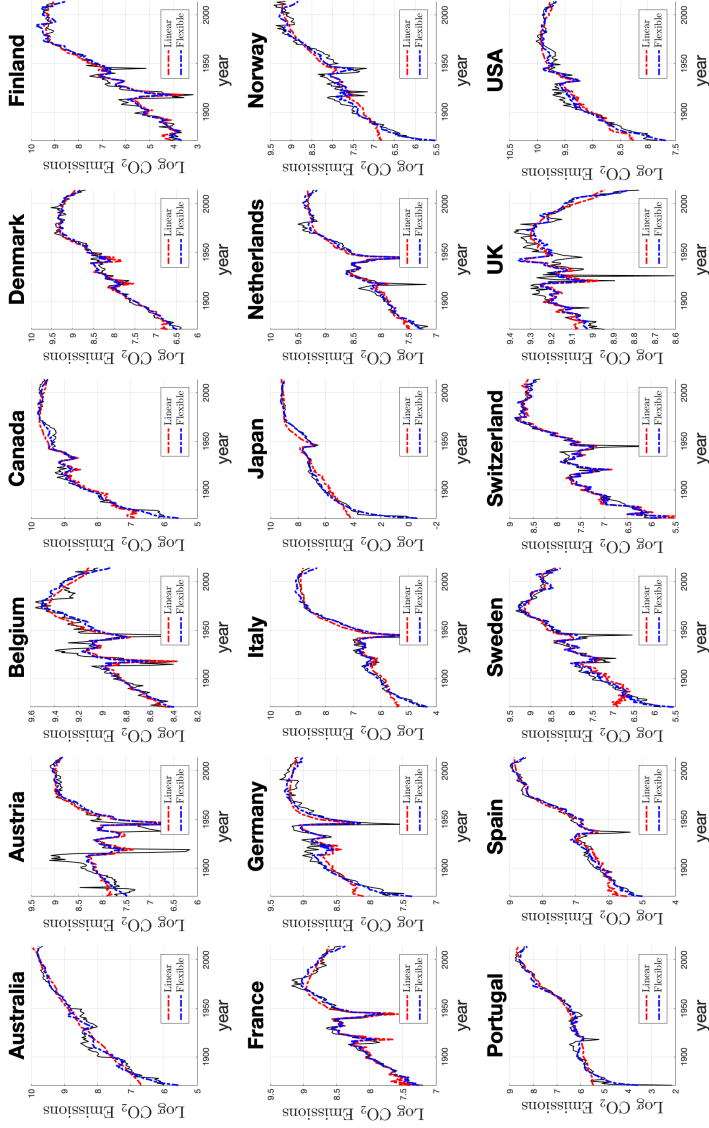


Figure 4.6: Estimation results for CO₂ emissions: actual values (black), fitted values under the CPR model $y_t = \tau_1 + \tau_2 t + \phi_1 x_t + \phi_2 x_t^2 + u_t$ (red), and fitted values under the GCPR model $y_t = \tau_1 + \tau_2 t + \tau_3 t^\theta + \phi_1 x_t + u_t$ (blue).

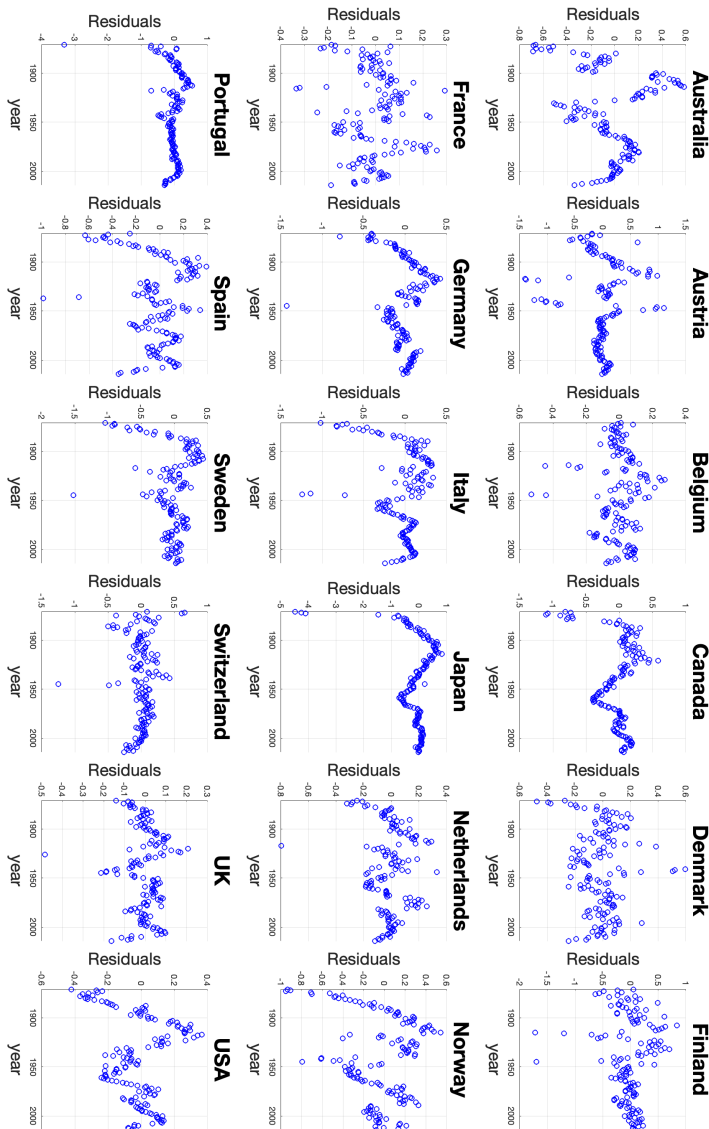


Figure 4.7: The residual series for each country under model specification (M1): $y_t = \tau_1 + \tau_2 t + \phi_1 x_t + \phi_2 x_t^2 + u_t$.

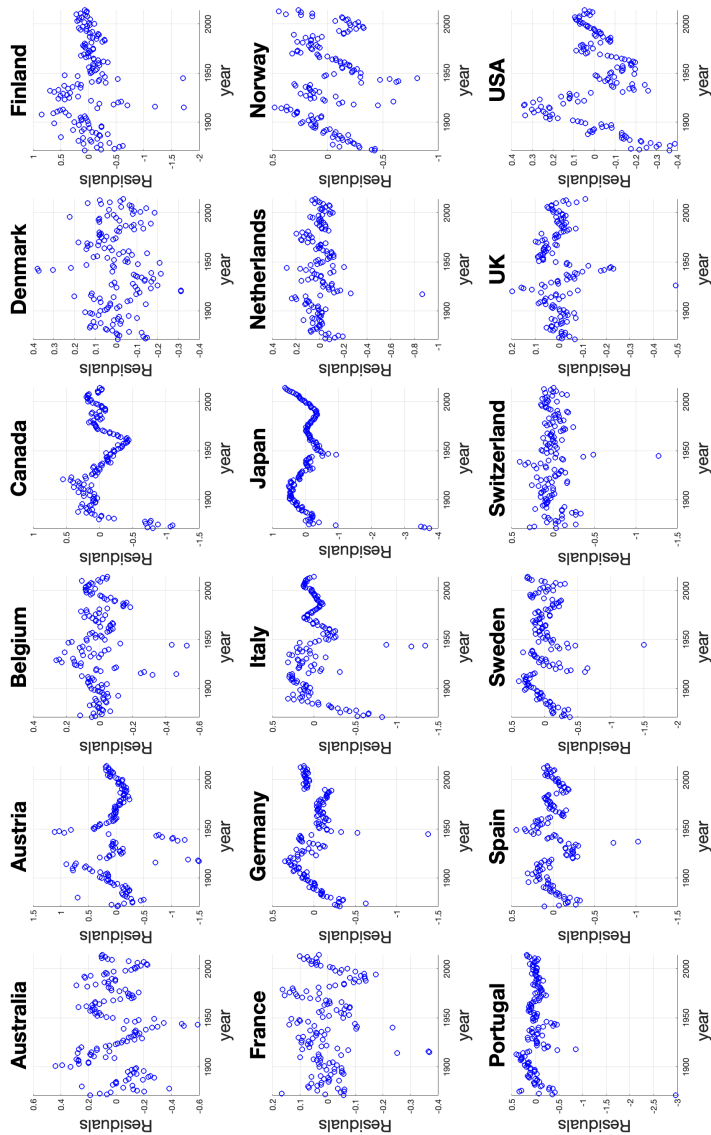


Figure 4.8: The residual series for each country under model specification (M2): $y_t = \tau_1 + \tau_2 t + \tau_3 t^2 + \phi_1 x_t + \phi_2 x_t^2 + u_t$.

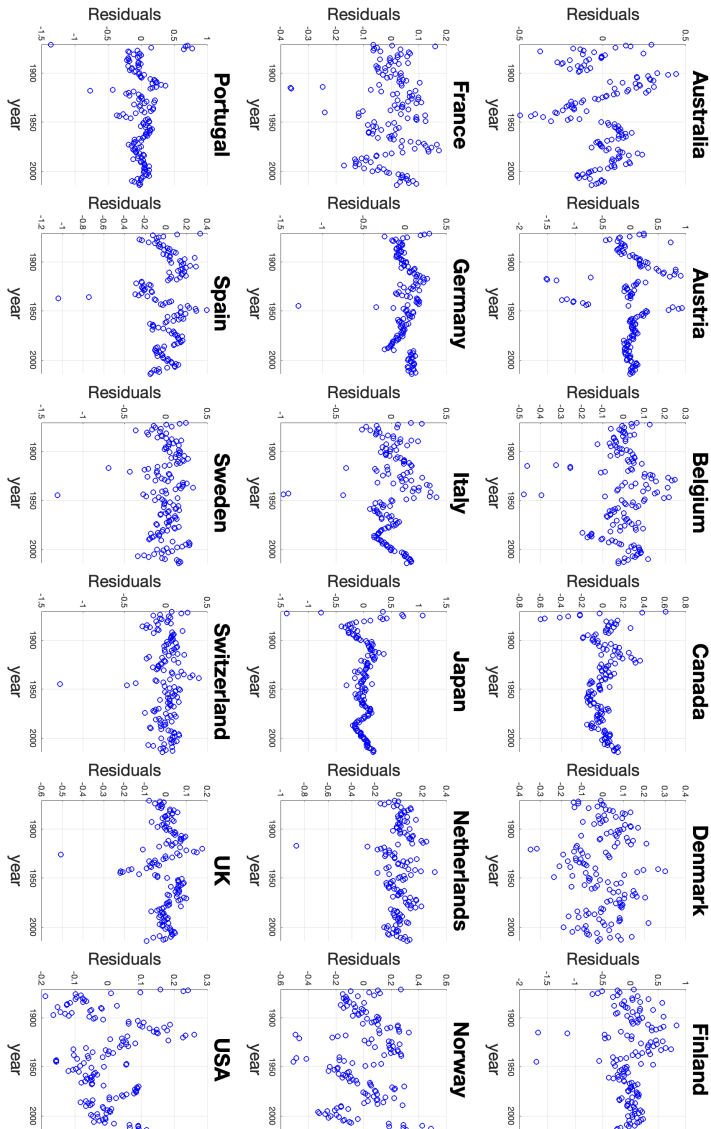


Figure 4.9: The residual series for each country under model specification (M3): $y_t = \tau_1 + \tau_2 t + \tau_3 t^{\theta} + \phi_1 x_t + \phi_2 x_t^2 + u_t$.

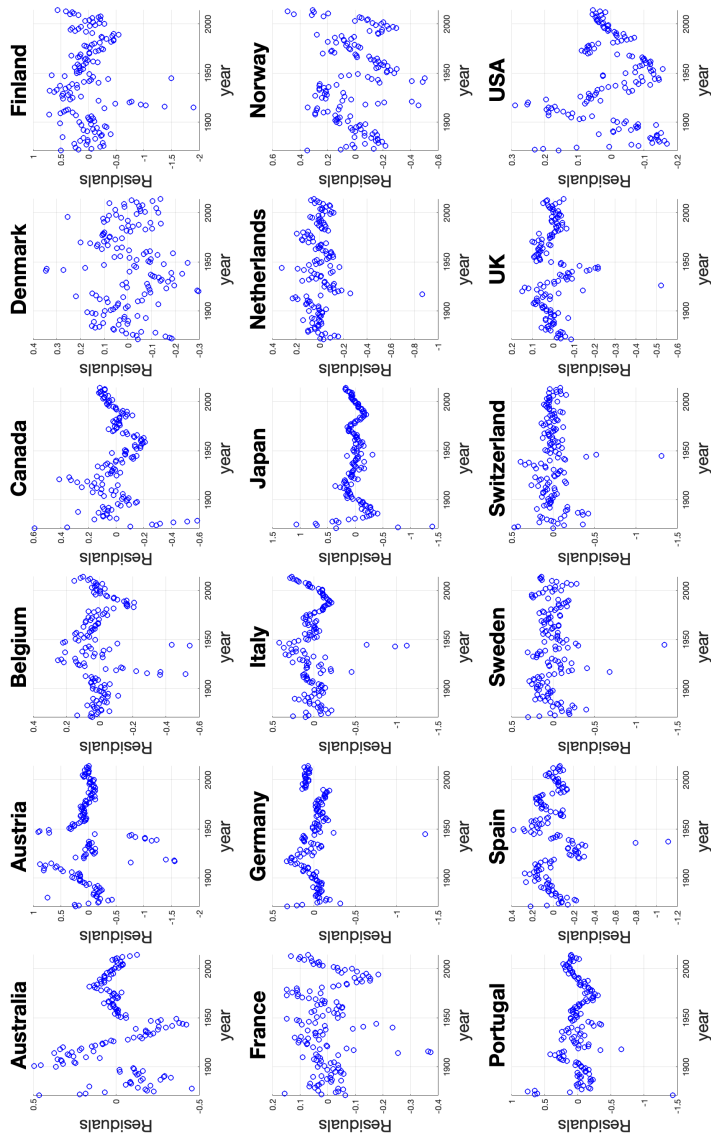


Figure 4.10: The residual series for each country under model specification (M4): $y_t = \tau_1 + \tau_2 t + \tau_3 t^\theta + \phi_1 x_t + u_t$.

5

Conclusion

This chapter provides an overall summary of the thesis. More detailed, method-specific conclusions can be found at the end of the core chapters, respectively. To understand the complexity of the real world, simple, interpretable nonlinear models are advantageous. We have developed different statistical frameworks to infer the finite sample behaviors of the estimators of unknown parameters. Nonlinearity, nonstationarity, and endogeneity are allowed in the models. We have evaluated the performances of the proposed methods through simulations and empirical applications. Practical recommendations are also made.

Chapter 2 studied models with structural breaks in level and/or trend slope. We have developed a feasible generalized least squares (FGLS) estimator of unknown break dates and derived the limiting distribution. The proposed methods improve the estimation accuracy compared to some existing methods in various scenarios. Based on our asymptotic results, we have also provided the new constructions of confidence sets/intervals of break dates by inverting Wald-type tests and multiple likelihood ratio tests. The estimation efficiency of the FGLS estimator is inherited by the confidence sets/intervals, thus improving both coverage and length in simulations.

We conducted inference on multivariate cointegrating polynomial regressions (CPRs) in Chapter 3. Parameters were constructed using a Fully Modified GLS approach. We tested for the null of nonlinear cointegration using subsampling approach and filtered residuals. Simulations revealed the advantages and disadvantages of different methods. We have found that all approaches underestimate parameter uncertainty. Therefore, confidence intervals are often too small. However, our method suffers the least from this problem.

In Chapter 4, we have extended the CPRs to allow possibly non-integer powers of deterministic trends. We have estimated the unknown powers and the parameters in the cointegrating relation jointly through nonlinear least squares. We have derived the joint limiting distribution for inference. This distribution is rather nonstandard and involves some second-order bias terms. The inference is therefore challenging. We then put forward a simulation-based method to conduct inference. Similarly, we have proved the validity of the common

subsampling test for nonlinear cointegration in our setting. A simulation study supported our theoretical results. Finally, we revisited the Environmental Kuznets Curve (EKC) hypothesis. We found that a flexible trend can fully capture the nonlinearity in the data thereby making higher order powers of log per capita GDP redundant. The resulting model was linear in log per capita GDP and suggested an alternative explanation for the recent slowdown in pollution. In contrast to the EKC hypothesis, our finding did not suggest that air quality will benefit from economic growth, but it could be affected by other factors such as technological progress.

This thesis contributes to the analysis of nonlinear, nonstationary time series, however, it is far from being comprehensive and complete. There are various unexplored research topics. For instance, apart from the preliminary study in Yang (2017), the statistical properties of break date estimators in models with multiple, unknown number of trend breaks remain largely unclear. Moreover, the current framework assumes (weak) stationarity of error processes. This assumption may not be fulfilled in practice. It is then interesting to extend the current method to allow $I(1)$ errors. A robust inference could be developed as in Harvey et al. (2007, 2009).

As for nonlinear cointegration, there are multiple possibilities. Consider a general parametric form $y_t = f(x_{t-1}, \gamma) + u_t$, where f is not necessarily polynomial. The statistical properties of this model have been well studied by Chan and Wang (2015). It is not hard to imagine that y_t may exhibit a seemingly stationary behavior for some function f , even though x_t is highly persistent. Such property leaves a room for studying predictive regression models. Moreover, current studies on nonlinear cointegration, except for CPRs, focus on univariate time series. For forecasting, it would be useful to consider multivariate time series or large panel models.

Another key issue is about test for nonlinear cointegration. It is well known that the usual subsampling test is highly conservative because of Bonferroni adjustments. It is also sensitive to the selection of parameters. To improve the performance, some simulation-based approaches such as bootstrap methods

can be helpful. To our best knowledge, the only available method is proposed by Hanck and Massing (2021). They have considered a heteroskedastic fixed regressor bootstrap and found good finite sample properties. Last but not least, in Chapter 4, we have found that the log per capita GDP may not solely explain the EKC. To obtain better insights and political implications, it is essential to break down the sources from the time factors.

“Well begun is half done,” as Aristotle said. Here is where we start the second half.

Bibliography

- Abadir, K. M. and J. R. Magnus (2005). *Matrix Algebra*. Cambridge University Press.
- Adams, R. A. and C. Essex (2016). *Calculus: A Complete Course*. Pearson Canada.
- Anderson, T. W. and D. A. Darling (1952). Asymptotic theory of certain “goodness of fit” criteria based on stochastic processes. *The Annals of Mathematical Statistics* 23, 193–212.
- Andrews, D. W. and X. Cheng (2012). Estimation and inference with weak, semi-strong, and strong identification. *Econometrica* 80, 2153–2211.
- Andrews, D. W. K. (1991). Heteroskedasticity and autocorrelation consistent covariance matrix estimation. *Econometrica* 59, 817–858.
- Andrews, D. W. K. and J. C. Monahan (1992). An improved heteroskedasticity and autocorrelation consistent covariance matrix estimator. *Econometrica* 60, 953–966.
- Baek, Y. I., S. J. Cho, and P. C. B. Phillips (2015). Testing linearity using power transforms of regressors. *Journal of Econometrics* 187, 376–384.
- Bai, J. (1994). Least squares estimation of a shift in linear processes. *Journal of Time Series Analysis* 15, 453–472.
- Bai, J. (1997). Estimation of a change point in multiple regression models. *Review of Economics and Statistics* 79, 551–563.
- Bai, J., R. L. Lumsdaine, and J. H. Stock (1998). Testing for and dating common breaks in multivariate time series. *Review of Economic Studies* 65, 395–432.

- Bai, J. and P. Perron (1998). Estimating and testing linear models with multiple structural changes. *Econometrica* 66, 47–78.
- Baily, M. N. (2002). The new economy: Post mortem or second wind? *Journal of Economic Perspectives* 16, 3–22.
- Baily, M. N. and R. Z. Lawrence (2001). Do we have a new E-conomy? *American Economic Review: Papers & Proceedings* 91, 308–312.
- Banerjee, A. V. and E. Duflo (2019). *Good Economics for Hard Times*. PublicAffairs.
- Bergamelli, M., A. Bianchi, L. Khalaf, and G. Urga (2019). Combining p-values to test for multiple structural breaks in cointegrated regressions. *Journal of Econometrics* 211, 461–482.
- Berk, K. N. (1974). Consistent autoregressive spectral estimates. *Annals of Statistics* 2, 489–502.
- Beutner, E., Y. Lin, and S. Smeeke (2021). GLS estimation and confidence sets for the date of a single break in models with trends. Working Paper.
- Bhattacharya, P. K. and P. J. Brockwell (1976). The minimum of an additive process with applications to signal estimation and storage theory. *Z. Wahrscheinlichkeitstheorie verw Gebiete* 37, 51–75.
- Bickel, P. J. and Y. R. Gel (2011). Banded regularization of autocovariance matrices in application to parameter estimation and forecasting of time series. *Journal of the Royal Statistical Society. Series B (Statistical Methodology)* 73, 711–728.
- Bickel, P. J. and E. Levina (2008). Regularized estimation of large covariance matrices. *Annals of Statistics* 36, 199–227.
- Black, S. E. and L. M. Lynch (2004). What’s driving the new economy?: The benefits of workplace innovation. *The Economic Journal* 114, F97–F116.

-
- Boden, T., G. Marland, and R. Andres (2017). Global, regional, and national fossil-fuel CO₂ emissions. Carbon Dioxide Information Analysis Center, Oak Ridge National Laboratory, U.S. Department of Energy, Oak Ridge, Tenn., U.S.A. http://cdiac.ess-dive.lbl.gov/trends/emis/tre_coun.html.
- Bolt, J., R. Inklaar, H. de Jong, and J. L. van Zanden (2018). Rebasing "maddison": New income comparisons and the shape of long-run economic development. <https://www.rug.nl/ggdc/historicaldevelopment/maddison/releases/maddison-project-database-2018>.
- Boot, T. and A. Pick (2020). Does modeling a structural break improve forecast accuracy? *Journal of Econometrics* 215, 35–59.
- Breitung, J. (2001). Rank tests for nonlinear cointegration. *Journal of Business & Economic Statistics* 19, 331–340.
- Brockwell, P. J. and R. A. Davis (1991). *Time Series: Theory and Methods* (2 ed.). Springer.
- Canjels, E. and M. W. Watson (1997). Estimating deterministic trends in the presence of serially correlated errors. *Review of Economics and Statistics* 79, 184–200.
- Carson, R. T. (2009). The Environmental Kuznets Curve: seeking empirical regularity and theoretical structure. *Review of Environmental Economics and Policy* 4, 3–23.
- Chan, N. and Q. Wang (2015). Nonlinear regressions with nonstationary time series. *Journal of Econometrics* 185, 182–195.
- Chang, S. Y. and P. Perron (2018). A comparison of alternative methods to construct confidence intervals for the estimate of a break date in linear regression models. *Econometric Reviews* 37, 577–601.

- Chang, Y., J. Y. Park, and P. C. B. Phillips (2001). Nonlinear econometric models with cointegrated and deterministically trending regressors. *The Econometrics Journal* 4, 1–36.
- Chang, Y., J. Y. Park, and P. C. B. Phillips (2004). Bootstrap unit root tests in panels with cross-sectional dependency. *Journal of Econometrics* 120, 263–293.
- Cheng, R. and M. Pourahmadi (1993). Baxter’s inequality and convergence of finite predictors of multivariate stochastic processes. *Probability Theory and Related Fields* 95, 115–124.
- Cheng, T.-C. F., C.-K. Ing, and S.-H. Yu (2015). Toward optimal model averaging in regression models with time series errors. *Journal of Econometrics* 189, 321–334.
- Cho, J. S. and P. C. B. Phillips (2018). Sequentially testing polynomial model hypotheses using power transforms of regressors. *Journal of Applied Econometrics* 33, 141–159.
- Choi, I. and P. Saikkonen (2010). Tests for nonlinear cointegration. *Econometric Theory* 26, 682–709.
- Dalla, V., L. Giraitis, and H. L. Koul (2014). Studentizing weighted sums of linear processes. *Journal of Time Series Analysis* 35, 151–172.
- Dasgupta, S., B. K. H. Wang, and D. Wheeler (2002). Confronting the environmental Kuznets curve. *Journal of Economic Perspectives* 16, 147–168.
- Davidson, J. (1994). *Stochastic Limit Theory*. Oxford University Press.
- Davidson, R. and J. G. MacKinnon (2004). *Econometric Theory and Methods*. Oxford University Press.
- Davies, R. B. (1977). Hypothesis testing when a nuisance parameter is present only under the alternative. *Biometrika* 64, 247–254.
- Davies, R. B. (1987). Hypothesis testing when a nuisance parameter is present only under the alternative. *Biometrika* 74, 33–43.

-
- de Jong, R. M. (2002). Nonlinear estimators with integrated regressors but without exogeneity. mimeo Michigan State University.
- Deng, A. and P. Perron (2006). A comparison of alternative asymptotic frameworks to analyse a structural change in a linear time trend. *Econometrics Journal* 9, 423–447.
- Dorfman, J. H. (1993). Should normality be a normal assumption? *Economics Letters* 42, 143–147.
- Dufour, J., A. Farhat, L. Gardiol, and L. Khalaf (1998). Simulation-based finite sample normality tests in linear regressions. *Econometrics Journal* 1, C154–C173.
- Dufour, J.-M. (2006). Monte carlo tests with nuisance parameters: A general approach to finite-sample inference and nonstandard asymptotics. *Journal of Econometrics* 133, 443–477.
- Dufour, J.-M. and L. Khalaf (2002). Simulation based finite and large sample tests in multivariate regressions. *Journal of Econometrics* 111, 303–322.
- Duggal, V. G., C. Saltzman, and L. R. Klein (1999). Infrastructure and productivity: A nonlinear approach. *Journal of Econometrics* 111, 47–74.
- Duggal, V. G., C. Saltzman, and L. R. Klein (2007). Infrastructure and productivity: An extension to private infrastructure and its productivity. *Journal of Econometrics* 140, 485–502.
- Elliott, G. and U. K. Müller (2007). Confidence sets for the date of a single break in linear time series regressions. *Journal of Econometrics* 141, 1196–1218.
- Eo, Y. and J. Morley (2015). Likelihood-ratio-based confidence sets for the timing of structural breaks. *Quantitative Economics* 6, 463–497.
- Estrada, F. and P. Perron (2017). Extracting and analyzing the warming trend in global and hemispheric temperatures. *Journal of Time Series Analysis* 38, 711–732.

- Feenstra, R. C., R. Inklaar, and M. P. Timmer (2015). The next generation of the penn world table. *American Economic Review* 105, 3150–3182.
- Findley, D. F. and C.-Z. Wei (1993). Moment bounds for deriving time series CLT's and model selection procedures. *Statistica Sinica* 3, 453–480.
- Friedrich, M., E. Beutner, H. Reuvers, S. Smeeke, J.-P. Urbain, W. Bader, B. Franco, B. Lejeune, and E. Mahieu (2020). A statistical analysis of time trends in atmospheric ethane. *Climatic Change* 162, 105–125.
- Gao, J., O. Linton, and B. Peng (2020). Inference on a semiparametric model with global power law and local nonparametric trends. *Econometric Theory* 36, 223–249.
- Gillingham, K. and J. H. Stock (2018). The cost of reducing greenhouse gas emissions. *Journal of Economic Perspectives* 32, 53–72.
- Gillingham, P. C. K. and W. Nordhaus (2018). Uncertainty in forecasts of long-run economic growth. *Proceedings of the National Academy of Sciences* 115, 5409–5414.
- Gordon, R. J. (2000). Does the “new economy” measure up to the great inventions of the past? *Journal of Economic Perspectives* 14, 49–74.
- Gordon, R. J. (2012). Is U.S. economic growth over? Faltering innovation confronts the six headwinds. *National Bureau of Economic Research*. NBER WP 18315.
- Gordon, R. J. (2014). The demise of U.S. economic growth: Restatement, rebuttal, and reflections. *National Bureau of Economic Research*. NBER WP 19895.
- Gordon, R. J. (2016). *The Rise and Fall of American Growth: The U.S. Standard of Living Since the Civil War*. Princeton University Press.
- Granger, C. W., T. Inoue, and N. Morin (1997). Nonlinear stochastic trends. *Journal of Econometrics* 81, 65–92.

-
- Greene, W. H. (2012). *Econometric Analysis (7th Edition)*. Pearson Education Limited.
- Grenander, U. and M. Rosenblatt (1957). *Statistical Analysis of Stationary Time Series*. Wiley.
- Grossman, G. M. and A. B. Krueger (1995). Economic growth and the environment. *The Quarterly Journal of Economics* 110, 353–377.
- Hamilton, J. D. (1994). *Time Series Analysis*. Princeton University Press.
- Hanck, C. and T. Massing (2021). Testing for nonlinear cointegration under heteroskedasticity. Available at <https://arxiv.org/abs/2102.08809>, arXiv:2102.08809v1.
- Hannan, E. and M. Deistler (2012). *The Statistical Theory of Linear Systems*. Society for Industrial and Applied Mathematics.
- Hansen, B. E. (1992). Consistent covariance matrix estimation for dependent heterogeneous processes. *Econometrica* 60, 967–972.
- Hansen, B. E. (2000). Sample splitting and threshold estimation. *Econometrica* 68, 575–603.
- Hansen, B. E. (2001). The new econometrics of structural change: Dating breaks in U.S. labor productivity. *Journal of Economic Perspectives* 15, 117–128.
- Harvey, D. I. and S. J. Leybourne (2014). Break date estimation for models with deterministic structural change. *Oxford Bulletin of Economics and Statistics* 76, 623–642.
- Harvey, D. I. and S. J. Leybourne (2015). Confidence sets for the date of a break in level and trend when the order of integration is unknown. *Journal of Econometrics* 184, 262–279.
- Harvey, D. I., S. J. Leybourne, and A. R. Taylor (2007). A simple, robust and powerful test of the trend hypothesis. *Journal of Econometrics* 141, 1302–1330.

- Harvey, D. I., S. J. Leybourne, and A. R. Taylor (2009). Simple, robust, and powerful tests of the breaking trend hypothesis. *Econometric Theory* 25, 995–1029.
- Hong, S. H. and P. C. B. Phillips (2010). Testing linearity in cointegrating relations with an application to purchasing power parity. *Journal of Business & Economic Statistics* 28, 96–114.
- Hu, Z., P. C. B. Phillips, and Q. Wang (2019). Nonlinear cointegrating power function regression with endogeneity. Working paper.
- Ing, C.-K., H.-T. Chiou, and M. Guo (2016). Estimation of inverse autocovariance matrices for long memory processes. *Bernoulli* 22, 1301–1330.
- Ing, C.-K. and C.-Z. Wei (2003). On same-realization prediction in an infinite-order autoregressive process. *Journal of Multivariate Analysis* 85, 130–155.
- Ing, C.-K. and C.-Z. Wei (2005). Order selection for same-realization predictions in autoregressive processes. *Annals of Statistics* 33, 2423–2474.
- Jansson, M. (2002). Consistent covariance matrix estimation for linear processes. *Econometric Theory* 18, 1449–1459.
- Jentsch, C. and D. N. Politis (2015). Covariance matrix estimation and linear process bootstrap for multivariate time series of possibly increasing dimension. *Annals of Statistics* 43, 1117–1140.
- Jiang, B., Y. Lu, and J. Y. Park (2019). Testing for stationary at high frequency. *Journal of Econometrics*. First published online 11 November 2019. <https://doi.org/10.1016/j.jeconom.2019.09.004>.
- Jones, C. I. (2002). Sources of U.S. economic growth in a world of ideas. *American Economic Review* 92, 220–239.
- Jones, C. I. and P. M. Romer (2010). The new Kaldor facts: ideas, institutions, population, and human capital. *American Economic Journal: Macroeconomics* 2, 224–245.

-
- Jorgenson, D. W., M. S. Ho, and K. J. Stiroh (2008). A retrospective look at the U.S. productivity growth resurgence. *Journal of Economic Perspectives* 22, 3–24.
- Jorgenson, D. W. and K. J. Stiroh (2000). Raising the speed limit: U.S. economic growth in the information age. *Brookings Papers on Economic Activity* 1, 125–211.
- Kang, X. and X. Deng (2020). An improved modified Cholesky decomposition method for precision matrix estimation. Available at <https://arxiv.org/abs/1710.05163>, arXiv:1710.05163v2.
- Kantz, H. and T. Schreiber (2003). *Nonlinear Time Series Analysis* (second ed.). Cambridge: Cambridge University Press.
- Kariya, T. and H. Kurata (2004). *Generalized Least Squares* (first ed.). John Wiley & Sons.
- Kim, C. and D. L. Zimmerman (2012). Unconstrained models for the covariance structure of multivariate longitudinal data. *Journal of Multivariate Analysis* 107, 104–118.
- Kim, J. and D. Pollard (1990). Cube root asymptotics. *Annals of Statistics* 18, 191–219.
- Klein, L. R., V. G. Duggal, and C. Saltzman (2004). Contributions of input-output analysis to the understanding of technological change: the information sector in the United States. In E. Dietzenbacher and M. L. Lahr (Eds.), *Wassily Leontief and Input-output Economics*, Chapter 17, pp. 311–336. Cambridge University Press.
- Kohli, P., T. P. Garcia, and M. Pourahmadi (2016). Modeling the cholesky factors of covariance matrices of multivariate longitudinal data. *Journal of Multivariate Analysis* 145, 87–100.
- Krugman, P. (1995). *The Age of Diminished Expectations*. MIT Press, Cambridge.

- Krugman, P. (1999). *The accidental theorist: And other dispatches from the dismal science*. W. W. Norton & Company.
- Kwiatkowski, D., P. C. B. Phillips, P. Schmidt, and Y. Shin (1992). Testing the null hypothesis of stationarity against the alternative of a unit root: How sure are we that economic time series have a unit root? *Journal of Econometrics* 54, 159–178.
- Lee, T.-H., H. White, and C. W.J.Granger (1993). Testing for neglected nonlinearity in time series models: A comparison of neural network methods and alternative tests. *Journal of Econometrics* 56, 269–290.
- Lewis, R. and G. C. Reinsel (1985). Prediction of multivariate time series by autoregressive model fitting. *Journal of Multivariate Analysis* 16, 393–411.
- Li, D., P. C. B. Phillips, and J. Gao (2020). Kernel-based inference in time-varying coefficient cointegrating regression. *Journal of Econometrics* 215, 607–632.
- Li, S. and O. Linton (2021). When will the Covid-19 pandemic peak? *Journal of Econometrics* 220, 130–157.
- Lin, Y. and H. Reuvers (2020a). Cointegrating polynomial regressions with power law trends: A new angle on the environmental kuznets curve. Available at <https://arxiv.org/abs/2009.02262>, arXiv:2009.02262v1.
- Lin, Y. and H. Reuvers (2020b). Efficient estimation by fully modified GLS with an application to the environmental kuznets curve. Available at <https://arxiv.org/abs/1908.02552>, arXiv:1908.02552v2.
- Lin, Y., Y. Tu, and Q. Yao (2020). Estimation for double-nonlinear cointegration. *Journal of Econometrics* 216, 175–191.
- Linton, O. and Q. Wang (2016). Nonparametric transformation regression with nonstationary data. *Econometric Theory* 32, 1–29.

-
- Liu, T., C. W. J. Granger, and W. P. Heller (1992). Using the correlation exponent to decide whether an economic series is chaotic. *Journal of Applied Econometrics* 7, S25–S39.
- Mark, N. C., M. Ogaki, and D. Sul (2005). Dynamic seemingly unrelated cointegrating regressions. *The Review of Economic Studies* 72, 797–820.
- McMurry, T. L. and D. N. Politis (2010). Banded and tapered estimates for autocovariance matrices and the linear process bootstrap. *Journal of Time Series Analysis* 31, 471–482.
- Moon, H. R. and B. Perron (2005). Efficient estimation of the seemingly unrelated regression cointegration model and testing for purchasing power parity. *Econometric Reviews* 23, 293–323.
- Newey, W. K. and K. D. West (1994). Automatic lag selection in covariance matrix estimation. *The Review of Economic Studies* 61, 631–653.
- Nordhaus, W. D. (2013). *The Climate Casino: Risk, Uncertainty, and Economics for a Warming World*. Yale University Press.
- Nordhaus, W. D. (2014). The perils of the learning model for modeling endogenous technological change. *The Energy Journal* 35.
- Nyblom, J. and A. Harvey (2000). Tests of common stochastic trends. *Econometric Theory* 16, 176–199.
- Panayotou, T. (1997). Demystifying the environmental kuznets curve: Turning a black box into a policy tool. *Environment and Development Economics* 2, 465–484.
- Park, J. Y. and P. C. B. Phillips (1988). Statistical inference in regressions with integrated processes: Part 1. *Econometric Theory* 4, 468–497.
- Park, J. Y. and P. C. B. Phillips (1999). Asymptotics for nonlinear transformations of integrated time series. *Econometric Theory* 15, 269–298.
- Park, J. Y. and P. C. B. Phillips (2001). Nonlinear regressions with integrated time series. *Econometrica* 69, 117–161.

- Perron, P. (1989). The great crash, the oil price shock, and the unit root hypothesis. *Econometrica* 57, 1361–1401.
- Perron, P. (2006). Dealing with structural breaks. In T. C. Mills and K. Patterson (Eds.), *Palgrave Handbook of Econometrics*, Volume 1, pp. 278–352. London: Palgrave Macmillan.
- Perron, P. and T. Yabu (2009). Estimating deterministic trends with an integrated or stationary noise component. *Journal of Econometrics* 151, 56–69.
- Perron, P. and X. Zhu (2005). Structural breaks with deterministic and stochastic trends. *Journal of Econometrics* 129, 65–119.
- Pesaran, M. H. and A. Timmermann (2005). Small sample properties of forecasts from autoregressive models under structural breaks. *Journal of Econometrics* 129, 183–217.
- Pesaran, M. H. and A. Timmermann (2007). Selection of estimation window in the presence of breaks. *Journal of Econometrics* 137, 134–161.
- Phillips, P. C. B. (1991). Optimal inference in cointegrated systems. *Econometrica* 59, 283–306.
- Phillips, P. C. B. (1995). Fully modified least squares and vector autoregression. *Econometrica* 63, 1023–1078.
- Phillips, P. C. B. (1998). New tools for understanding spurious regressions. *Econometrica* 66, 1299–1325.
- Phillips, P. C. B. (2007). Regression with slowly varying regressors and nonlinear trends. *Econometric Theory* 23, 557–614.
- Phillips, P. C. B. and B. E. Hansen (1990). Statistical inference in instrumental variables regression with I(1) processes. *The Review of Economic Studies* 57, 99–125.
- Phillips, P. C. B. and J. Y. Park (1988). Asymptotic equivalence of ordinary least squares and generalized least squares in regressions with integrated regressors. *Journal of the American Statistical Association* 83, 111–115.

-
- Phillips, P. C. B. and Z. Shi (2021). Boosting: Why you can use the hp filter. *International Economic Review*.
- Phillips, P. C. B. and V. Solo (1992). Asymptotics for linear processes. *Annals of Statistics* 20, 971–1001.
- Piaggio, M. and E. Padilla (2012). Co2 emissions and economic activity: Heterogeneity across countries and non-stationary series. *Energy Policy* 46, 370–381.
- Picard, D. (1985). Testing and estimating change-points in time series. *Advances in Applied Probability* 17, 841–867.
- Pourahmadi, M. (1999). Joint mean-covariance models with applications to longitudinal data: Unconstrained parameterisation. *Biometrika* 86, 677–690.
- Pourahmadi, M. (2013). *High-Dimensional Covariance Estimation*. John Wiley & Sons, Ltd.
- Prais, S. J. and C. B. Winsten (1954). Trend estimators and serial correlation. Cowles Foundation, Discussion Paper 383.
- Qu, Z. and P. Perron (2007). Estimating and testing structural changes in multivariate regressions. *Econometrica* 75, 459–502.
- Raj, B. and D. J. Slottje (1994). The trend behavior of alternative income inequality measures in the united states from 1947-1990 and the structural break. *Journal of Business & Economic Statistics* 12, 479–487.
- Richardson, M. and T. Smith (1993). A test for multivariate normality in stock returns. *The Journal of Business* 66, 295–321.
- Robinson, P. M. (2012). Inference on power law spatial trends. *Bernoulli* 18, 644–677.
- Romano, J. P. and M. Wolf (2001). Subsampling intervals in autoregressive models with linear time trend. *Econometrica* 69, 1283–1314.

- Rossi, B. (2021). Forecasting in the presence of instabilities: How do we know whether models predict well and how to improve them. *Journal of Economic Literature*, forthcoming.
- Saikkonen, P. (1992). Estimation and testing of cointegrated systems by an autoregressive approximation. *Econometric Theory* 8, 1–27.
- Shin, Y. (1994). A residual-based test of the null of cointegration against the alternative of no cointegration. *Econometric Theory* 10, 91–115.
- Siegmund, D. (1988). Confidence sets in change-point problems. *International Statistical Review* 56, 31–48.
- Sims, C. A., J. H. Stock, and M. W. Watson (1990). Inference in linear time series models with some unit roots. *Econometrica* 58, 113–144.
- Soong, T. T. (1973). *Random Differential Equations in Science and Engineering*. Academic Press, Inc.
- Stein, E. M. (1970). *Singular Integrals and Differentiability Properties of Functions*. Princeton University Press.
- Stern, D. I. (2004). The rise and fall of the environmental kuznets curve. *World Development* 32, 1419–1439.
- Stern, D. I. (2017). The environmental kuznets curve after 25 years. *Journal of Bioeconomics* 19, 7–28.
- Stypka, O., M. Wagner, P. Grabarczyk, and R. Kawka (2017). The asymptotic validity of “standard” fully modified OLS estimation and inference in cointegrating polynomial regressions. Working Paper.
- Tanaka, K. (2017). *Time Series Analysis: Nonstationary and Noninvertible Distribution Theory*. John Wiley & Sons.
- Tao, T. (2012). *Topics in Random Matrix Theory*. American Mathematical Society.

-
- Teräsvirta, T., D. Tjøstheim, and C. W. J. Granger (2010). *Modelling Nonlinear Economic Time Series*. Oxford University Press.
- Tjøstheim, D. (2020). Some notes on nonlinear cointegration: A partial review with some novel perspectives. *Econometric Reviews* 39, 655–673.
- van der Vaart, A. W. (1998). *Asymptotic Statistics*. Cambridge University Press.
- Vogelsang, T. J. and M. Wagner (2014). Integrated modified OLS estimation and fixed-b inference for cointegrating regressions. *Journal of Econometrics* 178, 741–760.
- Wagner, M. (2015). The Environmental Kuznets Curve, cointegration and nonlinearity. *Journal of Applied Econometrics* 30, 948–967.
- Wagner, M., P. Grabarczyk, and S. H. Hong (2020). Fully modified OLS estimation and inference for seemingly unrelated cointegrating polynomial regressions and the environmental Kuznets curve for carbon dioxide emissions. *Journal of Econometrics* 214, 216–255.
- Wagner, M. and S. H. Hong (2016). Cointegrating polynomial regressions: Fully modified OLS estimation and inference. *Econometric Theory* 32, 1289–1315.
- Wang, Q. (2015). *Limit Theorems for Nonlinear Cointegrating Regression*. World Scientific.
- Wang, Q. and P. C. B. Phillips (2009). Structural nonparametric cointegrating regression. *Econometrica* 77, 1901–1948.
- Wang, Q. and P. C. B. Phillips (2012). A specification test for nonlinear nonstationary models. *Annals of Statistics* 40, 727–758.
- Wang, Q., D. Wu, and K. Zhu (2018). Model checks for nonlinear cointegrating regression. *Journal of Econometrics* 207, 261–284.

- Wei, C.-Z. (1987). Adaptive prediction by least squares predictors in stochastic regression models with applications to time series. *Annals of Statistics* 15, 1667–1682.
- Wu, W. B. and M. Pourahmadi (2009). Banding sample autocovariance matrices of stationary processes. *Statistica Sinica* 19, 1755–1768.
- Xiao, H. and W. B. Wu (2012). Covariance matrix estimation for stationary time series. *The Annals of Statistics* 40, 466–493.
- Yamamoto, Y. (2018). A modified confidence set for the structural break date in linear regression models. *Econometric Reviews* 37, 974–999.
- Yang, J. (2017). Consistency of trend break point estimator with underspecified break number. *Econometrics* 5, 2225–1146.
- Zellner, A. (1962). An efficient method of estimating seemingly unrelated regressions and tests for aggregation bias. *Journal of the American Statistical Association* 57, 348–368.
- Zhang, T. and W. B. Wu (2015). Time-varying nonlinear regression models: Nonparametric estimation and model selection. *Annals of Statistics* 43, 741–768.

Impact Paragraph

In empirical time series analysis and forecasting, practitioners routinely adopt linear models as simple approximations to complicated, typically nonlinear economic relations. Linear models are useful in substantive economic problems. Ignoring nonlinear features, however, can be misleading for policymakers. The proposed methods in this thesis are additions to the toolkit for understanding nonlinear relations. In this chapter, I will discuss some practical relevance of the two main components of the dissertation – structural break inference and cointegrating polynomial regressions with a focus on the application of the Environmental Kuznets Curve (EKC) hypothesis.

As Hansen (2001)'s study lucidly remarked, “structural change is pervasive in economic time series relationships, and it can be quite perilous to ignore. Inferences about economic relationships can go astray, forecasts can be inaccurate, and policy recommendations can be misleading or worse.” However, as will be seen in the following debate, the break locations are not always evident to modelers. It suggests the necessity of (trend) break estimation and inference as considered in Chapter 2.

The debate mentioned above is about the changes in economic growth. In recent years, it has been widely acknowledged that the nature of world economy has been altered by some fundamental changes. The first of these changes is associated with the sustained growth in human capital throughout the world, which is summarized as one of the new Kaldor facts (Fact 5) by Jones and Romer (2010). For example, the years of schooling in advanced countries have a steady increase by 5 to 6 years in the past decades. The growth accounting exercise conducted by Jones (2002) shows that educational attainment has contributed around 30% of the growth in total output per hour in the United States during 1950 to 1993. As such, educational level is potentially an essential driving force of economic growth in the second half of the twentieth

century. Another change is the advancement of information technology (IT) including computers, semiconductors and software etc. Advocates of “New Economy” believe that the developments in IT, and the spillovers from IT into non-IT sectors can also drive the growth of economy persistently and stably; discussions can be found in, for instance, Jorgenson and Stiroh (2000), Baily and Lawrence (2001), Baily (2002), Black and Lynch (2004). The recent wave of technological innovation in machine learning, artificial intelligence, genetic engineering, 3D printing and so on, further leads many economists to see a bright future for economic growth.

On the other hand, skeptics argue that the growth rate cannot remain high permanently. At the end of the day, it must slow down because of a set of headwinds such as economic inequality (Gordon (2012, 2014)). At the center of debate for the U.S. economy, Robert J. Gordon takes the view that the demise of growth has already occurred four decades ago (Gordon (2014)). He enumerated some contradictions between the actual macro data of productivity growth and the predictions of techno-optimists. For example, the growth rate of output per hour is 1.59% on average from 1972 to 2013, substantially lower than the average 2.36% from 1891 to 1972. Earlier on, he found that the effects of New Economy on productivity growth are absent in 88% of the economy (Gordon (2000)). Not coincidentally, “You can see the computer age everywhere but in the productivity statistics”, Robert M. Solow has commented (cited in Jorgenson et al. (2008)). Krugman (1999) supported from a similar standpoint that “...the idea that we are living in an age of dramatic technological progress is mainly hype; the reality is that we live in a time when the fundamental things are actually not changing very rapidly at all”.¹ This economic growth debate shows opposite views of growth patterns, possibly leading to different policy proposals. For this reason, a formal statistical evaluation is thus of crucial importance. The results developed in Chapter 2 are potentially useful.

Similarly, the false conclusion of the EKC evidence may also bring an undesirable shift in the development policies of both developed and developing

¹ See the essay “Technology’s Wonders: Not So Wondrous”.

countries. The growth trajectories of the EKC support a “growing now, cleaning later” development tactic. It is because the EKC essentially implies an automatic reversing behavior of deterioration when the income is beyond a certain threshold. Without critical evaluation of the EKC, the implied strategy may lead to some irreversible changes in Earth’s climate system. According to Global Environment Outlook (GEO-6) published by the United Nations Environment Programme, “the environmental change sweeping the world is occurring at a faster pace than previously thought, making it imperative that governments act now to reverse the damage being done to the planet.” Moreover, it is a popular opinion that environmental change or deterioration can cause tremendous pressure on overall human well-being including economic prosperity and social justice. Therefore, believing the EKC evidence mistakenly may have further negative impact on ecosystems and thus human well-being. Chapters 3 and 4 consider the cointegrating polynomial models and offer some statistical methods that allow researchers to assess the EKC carefully.

All in all, the conclusions from this thesis are relevant for both academic and non-academic researchers. I sincerely hope that this thesis offers rigorous methods and valuable insights for policymakers to make the right decisions. With this thesis, I have contributed to the research community, but definitely, more general methods and models should be done in the future.

About the author

Yicong Lin was born on January 20, 1991, in Shenzhen, China. Between 2006 and 2009, he attended high school at Hongling Middle School in Shenzhen. He was the top scorer (149/150 points) of the college entrance examination (Gaokao) of mathematics in Guangdong Province in 2009. He continued his bachelor studies in Applied Mathematics (B.Sc.) with a specialization in Information and Computing Science and Economics (B.A.) at Shenzhen University during 2009-2013. During his bachelor's study, he received several top student's honors, for instance, the National Academic Scholarship awarded by the Ministry of Education of China. From February to June 2013, he became a volunteer teacher at the Primary School of Taxkorgan Tajik Autonomous County in Kashgar, Xinjiang Uygur Autonomous Region, China. In June 2013 he obtained both degrees and graduated with first-class honors. Subsequently, he received the High Potential Scholarship (full scholarship) and started his research master's study at Maastricht University. In April 2015, he participated the Econometric Game and became one of the winners. He obtained his M.Sc. degree in Economic and Financial Research with distinction (cum laude) from Maastricht University in August 2016.

After this graduation, Yicong started his Ph.D. under the supervision of Prof. Dr. Jean-Pierre Urbain and Dr. Stephan Smeekes. Prof. Dr. Eric Beutner became involved in the daily supervision in October 2016. In fall 2018, he spent three months at The Chinese University of Hong Kong as a visiting graduate student. This dissertation presents the findings of his research. Yicong presented his work at various international conferences. He is also a referee for the journals *Econometrics and Statistics*, *Economics Letters*. Since August 2020, he is a postdoctoral researcher at the Department of Econometrics and Data Science, Vrije Universiteit Amsterdam. At there, he teaches several courses and provides bachelor's and master's thesis supervision.